

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

INFORMATION, ONDELETTES, RÉSEAUX DE NEURONES, MÉTHODES
NUMÉRIQUES, ET MODÉLISATION ET PRÉDICTION DES SÉRIES
TEMPORELLES BOURSIÈRES : UNE ÉTUDE COMPARATIVE

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN INFORMATIQUE COGNITIVE

PAR

SALIM LAHMIRI

JUIN 2014

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je tiens d'abord à remercier mes directeurs de recherche, le professeur Mounir Boukadoum de l'UQÀM et le professeur Sylvain Chartier de l'université d'Ottawa, pour m'avoir conseillé et aidé à mener à terme ce projet d'étude. Merci pour tout l'encadrement et toute la patience.

J'aimerais également remercier les professeurs Esma Aimeur du département d'informatique et de recherche opérationnelle de l'université de Montréal, et Roger Villemaire et Albert Lejeune de l'UQÀM pour leurs précieux efforts pour évaluer cette thèse de doctorat et aussi pour leurs précieuses recommandations.

Un très grand merci à ma famille, mon père pour sa patience; mon adorable épouse et mes adorables filles, vous m'avez donné tant d'énergie.

TABLE DES MATIÈRES

RÉSUMÉ	
LISTE DES FIGURES.....	V
LISTE DES TABLEAUX.....	VII
INTRODUCTION	1
CHAPITRE I	
ÉTAT DE L'ART	1
1.1 Prévion des marchés boursiers.....	1
1.2 Finance comportementale et sentiment de l'investisseur	13
1.3 Le réseau de neurones artificiels spatio-temporel	17
1.4 Le chaos dans la prédiction boursière.....	18
1.5 la sélection des variables prédictives.....	20
1.6 Critique et proposition de la recherche.....	26
1.7 Contribution de la recherche.....	29
CHAPITRE II	
MÉTHODOLOGIE.....	34
2.1 Conception des expérimentations	34
2.2 La transformée en ondelette.....	40
2.2.1 La transformée ordinaire en ondelette discrète	40
2.2.2 Le choix du type d'ondelette	50
2.3 Le réseau de neurones artificiels.....	53
2.4 Les méthodes d'optimisation numérique.....	56
2.4.1 La méthode de la descente du gradient	58
2.4.2 La méthode du gradient conjugué	59
2.4.3 La méthode de Levenberg-Marquard	62
2.4.4 La détermination du pas de la future direction de recherche	63

2.5	Les machines à supports de vecteurs	65
2.5.1	Le cas des données linéairement séparables	65
2.5.2	Le cas des données non linéairement séparables et SVM non linéaire	68
2.6	Normalisation des données et mesures de performances	73
CHAPITRE III		
DONNÉES ET PRÉTRAITEMENTS		75
3.1	Les données	75
3.2	Processus ARMA et EGARCH	92
CHAPITRE IV		
LES RÉSULTATS		99
4.1	Résultats sans validation croisée	99
4.2	Résultats avec validation croisée	112
4.3	Conclusion partielle	117
CHAPITRE V		
DISCUSSION DES RÉSULTATS		119
CHAPITRE VI		
CONCLUSION		127
APPENDICE A		
LE CAPITAL ASSET PRICING MODEL		134
APPENDICE B		
MODÉLISATION DU PROCESSUS ARMA ET EGARCH		135
APPENDICE C		
LE TEST STATISTIQUE DE LA STATIONARITÉ		141
APPENDICE D		
LE TEST DE CAUSALITÉ DE GRANGER		144
APPENDICE E		
FONCTIONS D'AUTOCORRÉLATION ET D'AUTOCORRÉLATION		
PARTIELLE		147
APPENDICE F		
LA MATRICE HAC DE NEWEY-WEST		149
APPENDICE G		
TESTS D'AUTOCORRELATION DES RÉSIDUS		150
APPENDICE H		
LE TEST D'HETEROSKEDASTICITÉ ARCH-LM		152

BIBLIOGRAPHIE	154
---------------------	-----

LISTE DES FIGURES

Figure	Page
2.1	Systèmes de prédiction 39
2.2	Arbre de décomposition par ondelette ordinaire..... 41
2.3	Analyse par ondelette discrète d'un signal gaussien à pas d'échelle..... 44
2.4	Arbre de décomposition par transformée en paquet d'ondelette 46
2.5	Dé-bruitage d'un signal par la transformée en paquet d'ondelette 50
2.6	Les ondelettes Haar, Daubechies, et Symlet..... 52
2.7	Un exemple d'architecture de RNA..... 55
2.8	Problème de classification binaire et linéairement séparable 68
2.9	Séparation des données : fonction noyau linéaire versus polynomiale..... 71
3.1	Série temporelle du S&P500..... 78
3.2	Série temporelle $R(t)$ du S&P500 81
3.3	Histogramme et statistiques descriptives de $R(t)$ 81
3.4	Séries temporelles des variables économiques 83
3.5	Séries temporelles des indicateurs techniques 85
3.6	Séries temporelles des indicateurs techniques BR, MPM, et EMV 86
3.7	Séries temporelles des mesures de sentiment : ICI, VIX, Put-Call 88
3.8	Série temporelle de la volatilité du S&P500..... 89
3.9	Série temporelle du volume de transactions 90
3.10	Processus de la modélisation statistique de la série temporelle..... 92
3.11	La modélisation du processus ARMA(p,q) 93
3.12	La modélisation du processus EGARCH(m,n)..... 94

3.13	Prétraitement préalable au test de causalité de Granger	96
3.14	Schéma général du système prédictif.....	97
3.15	Signal de basses fréquences obtenu par AOD	97
3.16	Décomposition obtenue par APO	98
4.1	Résultats des simulations : variables économiques	103
4.2	Résultats des simulations : variables techniques	104
4.3	Résultats des simulations : approches historiques de la série	105
4.4	Résultats des simulations : variables de sentiment.....	106
4.5	Résultats des simulations : ensemble des informations	107
4.6	Résultats des simulations : analyse par ondelette discrète.....	108
4.7	Résultats des simulations : analyse par paquet d'ondelettes.....	109
4.8	Meilleures performances selon le type d'information/classifieur	110
4.9	Synthèse des résultats par type d'information	112
4.10	Résultats de la validation croisée.....	114
4.11	Effet du nombre de couches cachées sur la performance	116
D.1	Séries transformées et stationnaires	146

LISTE DES TABLEAUX

Tableau	Page
Aperçu sur la littérature : approche par indicateurs techniques	3
Aperçu sur la littérature : approche par variables économiques	5
Aperçu sur la littérature : approche par retards historiques	7
Techniques de sélection : approche de type filtre	23
Techniques de sélection : approche de type Wrapper	24
Techniques de sélection : approche de type embarqué (embedded).....	25
Critères de sélection de la meilleure décomposition par APO.....	48
Comparaisons des ondelettes Haar, Daubechies, et Symlet.....	52
Liste des variables économiques.....	82
Liste des indicateurs techniques.....	84
Variables prédictives sélectionnées par le test de causalité	100
Comparaison avec la littérature.....	126
Résultats de l'estimation du processus ARMA(1,1)	138
Résultats de l'estimation du processus AR(2)	139
Résultats de l'estimation du processus ARMA(2,0)-EGARCH(1,1).....	139
Résultats de l'estimation du processus ARMA(4,4)-EGARCH(1,1).....	140
Résultats du test ADF de stationarité	142
Résultats du test de stationarité pour les variables transformées	143
Détails des résultats du test de causalité de Granger.....	145

Résultats de la fonction d'autocorrélation	148
Résultat du test Breusch-Godfrey LM pour AR(2).....	151
Résultat du test Breusch-Godfrey LM pour ARMA(4,4)	151
Résultat du test ARCH LM pour ARMA(2,0)-EGARCH(1,1)	153
Résultat du test ARCH LM pour ARMA(4,4)-EGARCH(1,1)	153

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

AC	Fonction d'autocorrélation
AE	Algorithmes évolutionnaires
AIC	Akaike information criterion
AOD	Analyse en ondelette ordinaire et discrète
APO	Analyse en paquet d'ondelettes
ARCH	Auto Regressive Conditional Heterskedasticity
ARMA	Auto Regressive Moving Average
CS	Critère de sélection
EGARCH	Exponential general autoregressive conditional heteroskedastic
IA	Intelligence artificielle
ICI	Investor Confidence Index
LF	Logique floue
LM	Lagrange Multiplier (multiplicateur)
RNA	Réseau de neurones artificiels
SIC	Schwarz information criterion
SVM	Machines à supports de vecteurs
VIX	Volatility Index

RÉSUMÉ

L'intelligence artificielle connexionniste est devenue un outil de modélisation attrayant dans le contexte de la prédiction des séries temporelles boursières, grâce à la capacité des réseaux de neurones artificiels (RNA) de modéliser des séries temporelles bruitées ou incomplètes sans pour autant émettre des hypothèses fortes sur leurs distributions et la relation entre variables d'entrée et de sortie, contrairement aux modèles statistiques conventionnels. Par ailleurs, différentes catégories d'information peuvent servir d'entrées au système de prédiction. Ainsi plusieurs travaux empiriques montrent la possibilité de prévoir le marché boursier à partir d'information économique, technique ou historique (retards) tirée de séries temporelles boursières, ou de l'information extraite de l'analyse par ondelettes discrète (AOD). Pour réaliser la descente du gradient lors de l'entraînement du RNA, l'algorithme d'optimisation que l'on retrouve le plus souvent dans littérature sur la prédiction boursière est celui de Levenberg-Marquardt (L-M). Mais, il existe d'autres techniques d'optimisation plus avancées comme le Quasi-Newton, l'algorithme du gradient conjugué de type Polak-Ribière, Powell-Beale ou Fletcher-Reeves, et autres, et il n'a pas encore été établi quel l'algorithme donne les meilleures résultats.

L'objectif de la thèse vise à répondre aux questions suivantes :

Question.1 :

Quel type d'information d'entrée peut mener aux meilleures prédictions ?

Question.2 :

Une combinaison des différents types d'information peut-elle améliorer la qualité des prédictions ?

Question.3 :

Quel algorithme d'approximation numérique permet le meilleur apprentissage du RNA à retro-propagation d'erreur, et donc de donner les meilleurs résultats de prédiction ?

Question 4 :

Peut-on améliorer les résultats des variables prédictives traditionnelles en introduisant une dimension de traitement temps-fréquence obtenue par l'analyse multi-résolution; en l'occurrence l'analyse par paquets d'ondelettes (APO) ?

L'indice boursier américain S&P500, le plus utilisé par les investisseurs américains et internationaux, est retenu pour effectuer différentes simulations. L'objectif est de prédire ses hausses et baisses, et les mesures de performance considérées sont le taux de classification (prédiction) correcte, la sensibilité et la spécificité. Les machines à support de vecteurs (SVM) sont retenues pour être le modèle de référence vu leur capacité prédictive prouvée dans la littérature.

Les simulations ont montré les résultats suivants: (a) parmi les catégories d'informations classiques, les informations historiques et les informations sur l'état psychologique du marché sont les plus pertinentes pour la prédiction des tendances futures; (b) la combinaison des différentes catégories d'informations classiques (économique, technique, historique, sentiment) n'améliore pas la précision de la prédiction; (c) les différences entre les performances obtenues par différents algorithmes numériques pour l'entraînement du RNA sont minimales; cependant l'algorithme de Polak-Ribière semble être en général plus performant; (d) dans certains cas, l'algorithme L-M qui est le plus utilisé dans la littérature performe moins bien que les autres algorithmes considérés dans nos simulations; (e) contrairement à ce que l'on retrouve dans la littérature, le RNA performe mieux que le SVM; (f) l'information fréquentielle extraite par l'analyse multi-résolution par paquet d'ondelettes (APO) permet d'améliorer grandement la performance prédictive des RNA par rapport aux autres catégories d'information traditionnellement utilisées dans la littérature, et ce quelque soit l'algorithme numérique utilisé pour l'entraînement; (g) Finalement, l'usage de l'information extraite par APO permet d'obtenir une meilleure prédiction de la tendance future du marché S&P500 que l'information extraite par AOD utilisée dans la littérature. De ce fait, l'approche par l'APO est beaucoup plus simple à utiliser, contrairement aux approches traditionnelles qui requièrent beaucoup de prétraitements statistiques.

MOTS-CLÉS : réseaux de neurones, algorithmes numériques, machines à supports de vecteurs, information, ondelette, prédiction.

INTRODUCTION

Le marché boursier est une place d'échange sur laquelle se traitent des produits et instruments financiers (actions, obligations, options, contrats à termes, etc) avec le potentiel de réaliser de grands profits¹. Cependant, les études empiriques ont montré que prédire son état est une tâche très difficile à cause de la nature complexe, volatile et non-linéaire de ce marché (Bollerslev, 1987; Dunis et Zhou, 1998; Andersen et al., 1999; Andersen et al., 2003; Carnero et al., 2004; Koopman et al., 2005; Andersen et al., 2006; Hillebrand et Medeiros, 2006; Lieberman et Phillips, 2008). De plus, plusieurs facteurs peuvent affecter son comportement incluant les conditions économiques et des affaires, les événements politiques, les anticipations des intervenants sur le marché et leur état psychologique. De ce fait, on peut comparer le marché boursier à une fonction dynamique dont les données sont volatiles, irrégulières et bruitées; d'où l'énorme difficulté de modéliser les séries temporelles (chronologiques) boursières à des fins de prédictions. En général, la modélisation de ces séries exige la collecte de données, l'analyse des caractéristiques et leurs relations avec la série en question, et finalement la production d'un système qui facilite la prédiction des événements ou des situations futures. Deux familles de méthodes sont utilisées pour modéliser les séries temporelles boursières: les méthodes statistiques et les méthodes de l'intelligence artificielle (IA).

¹ Dans cette thèse nous faisons référence au profit (ou rendement) comme étant la différence entre le prix d'achat et le prix de vente d'un même titre (action) boursier. Nous ne sommes pas concernés par le profit (rendement) excédentaire (Fama et French, 2002; Goyal et Welch, 2003, 2008) calculé par la différence entre le profit réalisé par un titre boursier et le taux de rendement d'un titre sans risque (par exemple un titre du bon du trésor). Comme nous allons le mentionner au niveau du chapitre 1, l'objectif de notre thèse est de calculer le profit (rendement) afin de déterminer le signe de la tendance future du marché que nous voulons modéliser et prédire.

Les méthodes statistiques sont traditionnellement les plus utilisées; notamment les processus ARIMA (Autoregressive Integrated Moving Average; voir Box et Jenkins, 1976; Hamilton, 1989; Hamilton, 1994). Elles sont fondées sur des formalismes mathématiques et offrent la possibilité de faire de l'inférence statistique en plus de la prédiction (Brockwell et Davis, 2002; Wooldridge, 2002; Terence et Mills, 1999; Harris et Sollis, 2010). Cependant, les méthodes statistiques conventionnelles sont contraintes par des postulats comme la normalité des distributions qui sont difficiles à rencontrer dans les données réelles. En outre, ces méthodes sont pour la plupart basées sur des modèles linéaires et donc ont de la difficulté à capturer la dynamique des séries lorsque ces dernières sous tendent un processus non-linéaire.

Par exemple, les modèles ARIMA ne sont pas capables de capturer l'irrégularité des séries temporelles boursières, car ils supposent que les valeurs prochaines de la série temporelle ont une relation linéaire avec les valeurs présentes et passées.

Plusieurs méthodes de l'IA comme les réseaux de neurones artificiels (RNA) (Enke et Thawornwong, 2005; Huang et al. 2007; de Faria et al., 2009), la logique floue (LF) (Atsalakis et Valavanis, 2009; Boyacioglu et Avci, 2010; Esfahanipour et Aghamiri, 2010), les algorithmes évolutionnaires (AE) (Tan et al., 2005; Kim, 2006; Majhi et al. 2009) ont été proposées comme alternatives aux modèles statistiques pour modéliser les séries temporelles et pour améliorer la qualité des prédictions. Les méthodes de l'IA sont robustes à différents types de bruit et s'adaptent facilement aux problèmes non-linéaires ou mal définis. Les RNA² - dont la structure de base et le mécanisme de fonctionnement s'inspirent du fonctionnement du système neuronal biologique - sont récemment devenus les plus utilisés dans le domaine de la prédiction boursière (Atsalakis et Valavanis, 2009), cela pour plusieurs raisons : Premièrement, les RNA sont des modèles non-paramétriques bien adaptés à la modélisation des séries temporelles car ils ne supposent aucun postulat relatif au processus qui génère les

² Le terme RNA sera utilisé au singulier et au pluriel dans le texte.

données ou la nature de leur distribution (Zhang, 1998). Étant donné que les réseaux de neurones ne sont pas utilisés à des fins d'inférence, ils sont moins susceptibles au problème de mauvaise spécification que les modèles statistiques paramétriques. Deuxièmement, ils sont capables de réaliser une adéquation entre les entrées et les sorties même lorsque la vraie nature de la relation entrée-sortie est inconnue ou est difficile à formuler mathématiquement. En effet, un RNA est capable d'approximer n'importe quelle fonction mesurable et continue avec une erreur arbitrairement petite (Cybenko, 1989; Hornik, 1991). Troisièmement, le réseau de neurones artificiel peut optimiser ses paramètres internes par une technique d'entraînement par l'exemple. Par ailleurs, il peut demeurer performant, stable et robuste dans un environnement bruyé (Khashei et Bijari, 2010). Quatrièmement, le RNA utilise moins de paramètres qu'un modèle de forme polynomiale, S-pline ou expansion trigonométrique pour obtenir le même taux d'approximation (Chakradhara et Narasimhan, 2007).

L'apprentissage des RNA pour pouvoir approximer une fonction quelconque nécessite l'adoption d'un algorithme prédéterminé. Il en existe plusieurs, le plus utilisé dans la littérature (Atsalakis et Valavanis, 2009; Wang et al., 2011) étant la descente du gradient, en particulier l'algorithme de Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1966; Haykin, 2001; Nocedal et Wright, 2006; Eligius et Boglárka, 2010).

Le choix des variables prédictives est d'une grande importance pour la modélisation et la prédiction des séries temporelles boursières, car il détermine l'interprétation des résultats et le niveau de performance des prévisions (John et al., 1994; Kohavi et Sommerfield, 1995; Cost et Salzberg, 1996; Koller et Sahami, 1996; Liu et Yu, 2005).

Il existe en général quatre approches – ou écoles de pensées – pour le choix des variables prédictives : l'analyse fondamentale, l'analyse technique, l'approche comportementale, et l'analyse statistique des rendements boursiers passés. En

d'autres mots, chacune de ces approches³ utilise une catégorie d'information spécifique.

L'analyse fondamentale est basée sur la théorie de l'économie financière. Selon cette théorie, la variation du prix est la conséquence de l'arrivée de nouvelles informations économiques affectant le mécanisme de formation de prix. Par conséquent, l'analyse fondamentale (Fabozzi, 2001; Benjamin et David, 2004) est l'étude des données et des rapports sur l'économie qui sont publiés à des intervalles réguliers (quotidien, mensuel ou trimestriel) selon la nature de l'information. Les différents aspects de l'économie incluent - mais n'y sont pas limités- les taux d'intérêts, les taux de changes, l'inflation, l'emploi, la productivité, etc. L'état actuel de l'économie est très important car le marché qui est composé des acteurs économiques et des investisseurs répond aux anticipations économiques. Par exemple, si les anticipations des agents économiques et des investisseurs sont consistantes avec l'état actuel de l'économie, alors le marché pourrait ne pas réagir à ces informations. Par contre, si les anticipations du marché ont été significativement supérieures ou inférieures aux informations actuelles; alors il devrait vraisemblablement réagir. Plus concrètement, supposons que les investisseurs s'attendaient à une augmentation du produit intérieur brut (PIB) mais que les rapports statistiques du gouvernement annoncent plutôt une baisse dans le PIB. Cette information veut dire implicitement que les compagnies cotées en bourse vont réaliser moins de profit. Alors, les investisseurs vont vendre leurs titres sur le marché boursier ce qui va engendrer une chute dans l'indice boursier, dûe à une baisse généralisée des prix des entreprises cotées sur le marché en question.

³ Une classification (taxonomie) des méthodes de sélection des variables prédictives est présentée au niveau du chapitre qui traite de l'état de l'art : section 5 du chapitre I.

En somme, l'information macroéconomique affecte les rendements boursiers à travers son impact sur les prix boursiers et ; donc influence également la tendance du marché boursier vers la hausse ou vers la baisse.

L'avantage de l'analyse fondamentale est qu'elle est fondée sur le concept du prix d'équilibre : le prix d'un actif dans un marché efficient retourne toujours à sa vraie valeur fondamentale. Ceci est un concept théorique fondamental en économie financière. Pour déterminer le prix d'équilibre, l'analyse fondamentale utilise de l'information concernant les entreprises (activité/secteur/comptabilité...), les marchés et l'économie. Plus concrètement, le prix d'équilibre d'un titre boursier est calculé à partir des flux monétaires qu'il va engendrer dans le futur. En d'autres mots, c'est la valeur actualisée de son prix et des dividendes qu'il va verser dans le futur. Le principal désavantage de l'analyse fondamentale est que la réception et le traitement de l'information (par exemple le niveau actuel du PIB) prend beaucoup de temps (de quelques jours à quelques semaines) pour pouvoir l'exploiter rapidement en termes de décision sur le marché. Cependant, il faut noter que l'analyse fondamentale contemporaine peut utiliser de l'information continue en temps réel comme les taux de changes, les taux d'intérêts, le prix de l'or, le prix du pétrole, et les niveaux des indices boursiers des autres place financières à travers le monde. Mais pour certains indicateurs économiques comme le produit intérieur brut, le taux de chômage, l'inflation, la consommation et la production industrielle; l'information macroéconomique ne peut pas être implémentée en temps réel dans le cadre l'analyse fondamentale contemporaine dans la mesure où ces informations (facteurs économiques cités précédemment) ne sont rendues publiques qu'à chaque 3 à 4 mois. Ceci est dû au fait qu'elles requièrent de la collecte d'information au niveau national, du traitement, et de la compilation. Finalement, un dernier désavantage de l'analyse fondamentale est le fait qu'elle n'inclut pas les aspects psychologiques qui influencent le marché.

L'analyse technique (Reilly et Brown, 1997; Schwager, 1999), contrairement à l'analyse fondamentale, n'utilise pas l'information économique pour prédire le mouvement du prix; mais l'information technique comme son nom l'indique. Elle est basée sur l'étude de l'historique de l'indice boursier et des indicateurs de l'activité du marché boursier. Par exemple, elle requiert l'utilisation de graphes de tendances et des indicateurs pour la prédiction du mouvement du prix. En effet, les graphiques montrent comment le prix varie, le moment de sa tendance, et la force de sa tendance (Reilly et Brown, 1997; Schwager, 1999).

Par exemple, l'oscillation et le moment (momentum) décrivent bien l'action du marché (Reilly et Brown, 1997; Schwager, 1999). De ce fait, l'analyse technique utilise de l'information passée et contemporaine du prix pour prédire sa valeur ou sa tendance future.

Il existe plusieurs avantages à ce type d'analyse. Premièrement, le graphique permet de détecter de façon plus ou moins claire et facile la tendance du prix (Levich et Thomas, 1993; Blume et al., 1994; Neely et al., 1997). Ceci est dû au fait que le prix continu vraisemblablement sur sa tendance pour une certaine période. Deuxièmement, le graphique permet de détecter les caractéristiques du marché et ; par conséquent; de prédire le mouvement du prix. Ceci est dû au fait que le cycle du marché est répétitif dans certaines périodes ce qui permet de réaliser d'énormes profits (Taylor et Allen, 1992; Brailsford et Faff, 1993; McKenzie, 2001). Troisièmement, les graphes et les indicateurs techniques fournissent rapidement un énorme ensemble d'information. Il y a plus de cinquante indicateurs qui renseignent sur les niveaux de support, de résistance, de moment, de volatilité, et des caractéristiques des transactions. Finalement, l'analyse technique ne requière pas beaucoup de temps pour être implémentée à cause de la puissance des ordinateurs contemporains, et elle est moins coûteuse que l'analyse fondamentale en termes de collecte et de traitement des données.

L'inconvénient majeur de l'analyse technique est le fait qu'elle permet de tirer profit juste au début de la nouvelle tendance du marché et donc sur une toute petite courte durée. En effet, le marché boursier peut changer brusquement et sans "préavis" de direction; or l'analyse technique permet de tirer profit seulement lorsque le marché est cyclique. Finalement, l'analyse technique n'est pas fondée sur un concept théorique.

L'analyse statistique de l'historique des rendements est fondée sur l'analyse de la série temporelle pour détecter une éventuelle mémoire. En d'autres mots, il s'agit de la modélisation de la dynamique interne de la série en question en se basant sur l'information extraite de ses caractéristiques statistiques. En effet, la question de la présence d'une mémoire longue/courte et stochastique dans les rendements boursiers avait été depuis longtemps l'un des sujets les plus étudiés (Cheung et Lai, 1995; Grang et et Hyung, 2004; Mukherjee et al., 2011; Rege et Martin, 2011) tant au niveau théorique qu'empirique. Par exemple, si les rendements boursiers ont une mémoire longue - une dépendance à long terme - alors ils illustrent une auto-corrélation positive et significative entre les observations (Geweke et Porter-Hudak, 1983; Lo, 1991; Crato, 1994; Cheung et Lai, 1995).

Donc, le fait que les séries des rendements ne sont pas indépendantes dans le temps suggère d'analyser leur historique afin de prédire leurs tendances futures; et ainsi réaliser des profits spéculatifs consistants. En d'autres mots, les rendements futurs sont prévisibles étant donné une connaissance à priori des rendements passés.

Techniquement parlant, les séries temporelles boursières qui montrent une forte dépendance - existence d'une mémoire longue - ont une fonction d'auto-corrélation qui diminue lentement dans le temps. En d'autres mots, la fonction d'auto-corrélation permet de détecter combien de retards influencent la valeur actuelle de la série. Par exemple, si P est le niveau actuel de l'indice boursier au temps t , alors la fonction d'auto-corrélation permet de détecter combien de valeurs passées ($P(t-1)$, $P(t-2)$, $P(t-$

3),...) affecte le prix actuel $P(t)$. Ainsi, la mémoire dans le contexte d'une série temporelle représente les observations passées de ladite série qui sont corrélées avec l'observation actuelle (présente). Ceci nécessite une certaine classe de modèles pour décrire ce type de comportement. La classe la plus populaire (Granger and Joyeux, 1980; Hosking, 1981; Robinson, 2003) est celle qui adopte les processus ARMA ou ARFIMA qui sont des modèles statistiques basés sur des processus linéaires. Par exemple, le processus ARMA permet de modéliser la mémoire au niveau du terme auto-regressif (AR) et d'effectuer un lissage à long terme de la série en question par le terme moyenne mobile (MA : Moving Average).

Le principal avantage de cette approche est qu'il est facile d'analyser la série temporelle et de détecter sa mémoire longue si elle existe. De plus, elle est facile à implémenter et à modéliser en un peu de temps. En fait, c'est l'approche la plus facile et la plus rapide car elle ne requiert pas autant d'informations que nécessitent les autres approches comme l'analyse fondamentale, l'analyse technique ou l'approche de la finance comportementale que nous allons introduire juste après. Par contre, le principal inconvénient de l'approche de l'analyse statistique est qu'elle ignore toute influence des autres facteurs économiques, marchés ou psychologiques.

L'approche comportementale (Herbert, 1987; Barberis et al., 1998; Kahneman and Riepe; 1998, Shapira and Venezia, 2000; Shefrin, 2002) (ou finance comportementale) est fondée sur l'application des principes psychologiques pour étudier et comprendre le processus de la décision financière. De ce fait, elle utilise de l'information psychologique pour expliquer et prédire le marché boursier. Elle considère la finance traditionnelle (exemple; l'analyse fondamentale) comme incomplète sur le fait qu'il existe certaines anomalies du marché qui ne sont pas expliquées par la théorie financière et économique.

Parmi ces anomalies (Hirshleifer, 2001), on peut citer l'ancrage⁴, l'heuristique⁵ et l'excès⁶ de confiance. En effet, l'investisseur semble ne pas réagir de façon rationnelle face aux nouvelles informations économiques et financières. Ainsi, la finance comportementale introduit des aspects « behavioristes » dans le processus de prise de décision. L'objectif est de comprendre la réflexion chez l'investisseur pour prévoir ses décisions. En d'autres mots, la finance comportementale introduit des biais psychologiques afin d'améliorer la prédiction du marché.

L'avantage majeur de la finance comportementale est qu'elle met l'accent sur le fonctionnement de l'esprit humain en général et le profil psychologique de l'investisseur pour expliquer pourquoi et comment le marché boursier est inefficace; ce qui constitue par conséquent une opportunité pour réaliser des profits. En particulier, la finance comportementale contribue à l'explication de l'inefficacité du marché boursier en apportant de nouvelles théories et observations d'ordre psychologique. Alors, il s'agit d'exploiter ces éléments pour modéliser les mouvements du marché boursier et de le prédire.

Le désavantage majeur de la finance comportementale est justement le fait qu'elle n'est pas fondée sur la théorie financière et économique. En d'autres mots, elle n'offre pas des arguments objectifs. De plus, elle s'avère être difficile à implémenter.

Nous avons brièvement passé en revue les catégories d'informations qui sont couramment utilisées pour modéliser et prédire les séries temporelles boursières,

⁴ L'ancre est le prix actuel le plus reconnu. En absence d'une meilleure information, l'historique des cours est déterminant de l'ancre. Les investisseurs choisissent les ancres selon leur opinion. Après avoir reçu les nouvelles informations, les investisseurs ajustent les ancres passés.

⁵ L'heuristique est le processus qui mène l'investisseur à sa décision, soit par tâtonnement soit par erreur. Selon l'approche comportementale, l'heuristique est une forme où l'investisseur cherche à avoir des décisions identiques à celles des autres.

⁶ L'excès de confiance est une surestimation de la capacité de l'investisseur. Il particulièrement constaté chez l'investisseur expérimenté.

chacune d'elles présentant des avantages et inconvénients par rapports aux autres. La question qui se pose alors est : existe-t-il une autre technique qui fournit une information mieux en mesure de caractériser les mouvements du marché boursier et donc de mieux le prédire ? En d'autres mots, existe-t-il une technique permettant d'extraire des caractéristiques internes et spécifiques d'une série temporelle? Une telle information pourrait être bénéfique pour la modéliser et la prédire. L'analyse multi-résolution représente une approche prometteuse dans ce sens. En particulier, l'analyse multi-résolution par ondelettes (Percival et Andrew, 2000; Jensen et la Cour-Harbo, 2001) est un outil de traitement du signal largement utilisé dans les corps de génie. Elle consiste à décomposer mathématiquement un signal par un ensemble de fonctions de base appelées ondelettes qui varient en fréquence et en temps. Donc, l'ondelette permet une localisation en fréquence et une localisation temporelle du signal d'origine. Il en découle des signaux de basses fréquences et des signaux de hautes fréquences pour caractériser le signal initial, permettant ainsi de décomposer une série temporelle boursière en une composante qui reflète sa tendance à long terme et une autre qui reflète sa variation locale. Ces deux composantes prises conjointement peuvent être très bénéfiques pour entraîner un RNA ou SVM. Malheureusement, l'analyse multi-résolution demeure inconnue dans le milieu de la finance et rarement employée pour la modélisation des séries temporelles boursières comme nous allons le voir dans le chapitre qui porte sur la revue de littérature.

Rappelons que le réseau de neurones artificiels (RNA) est devenu l'outil d'intelligence artificielle le plus utilisé pour modéliser les séries temporelles boursières grâce à sa capacité à modéliser des séries temporelles bruitées ou incomplètes sans pour autant émettre des hypothèses fortes sur leurs distributions et la relation entre variables d'entrée et de sortie, contrairement aux modèles statistiques conventionnels (Atsalakis et Valavanis, 2009; Khashei et Bijari, 2010). De ce fait, l'objectif de cette thèse est de déterminer la catégorie d'information d'entrée et le type d'algorithme d'approximation numérique qui permettent le meilleur

apprentissage du RNA de type perceptron multicouches⁷ à retro-propagation d'erreur, et donc de donner les meilleurs résultats de prédiction.

La contribution au niveau informatique est d'étudier l'impact du choix du type d'information et de l'algorithme d'apprentissage sur la capacité prédictive du RNA à rétro-propagation d'erreur. Au niveau cognitif, nous utilisons l'état psychologique du marché comme entré au RNA. L'état psychologique du marché est représenté par des mesures de sentiment de l'investisseur tel que proposé par la finance comportementale.

De plus, nous identifions les intrants qui montrent un lien causal plus fort avec la tendance future du marché (test de causalité de Granger), pour retenir seulement les variables qui influencent la décision de l'investisseur. Par conséquent, ces variables sélectionnées pourraient améliorer la qualité des prédictions.

L'indice boursier américain S&P500, le plus utilisé par les investisseurs américains et internationaux, est retenu pour effectuer les différentes simulations. L'objectif est de prédire sa tendance future à la hausse ou à la baisse.

Nous nous attendons à ce que l'information obtenue par l'analyse multi-résolution et le RNA entraîné par un algorithme numérique de type descente de gradient conjugué (qui utilise une direction de recherche du minimum à chaque itération) permettrait de fournir une meilleure prédiction que les autres informations traditionnelles et l'algorithme L-M standard.

Rappelons trois points importants. Le premier est que les informations macroéconomiques, techniques, historiques et psychologiques affectent les rendements boursiers à travers leurs impacts sur les prix boursiers. Par conséquent,

⁷ Le perceptron multicouche est le type de RNA adopté dans notre thèse. De ce fait, le terme RNA est utilisé dans la thèse pour indiquer l'utilisation du perceptron multicouches.

elles influencent également les tendances haussières et baissières du marché boursier. Le deuxième point est que nous faisons de la prédiction de la tendance future du marché boursier : une hausse ou une baisse éventuelle. Donc, il s'agit d'un problème de classification appliqué à une série temporelle. De ce fait, nous ne faisons pas de la prédiction du rendement excédentaire du marché boursier par rapport à un taux d'intérêt sans risque. Le troisième point est que nous ne faisons pas une comparaison entre modèles statistiques et modèles de réseaux de neurones artificiels. Nous utilisons les techniques statistiques pour prétraiter les données et d'y extraire les informations pertinentes à la modélisation par réseaux de neurones.

En particulier, les méthodes statistiques seront employées pour rendre les données stationnaires, modéliser le processus des rendements et extraire sa volatilité, et finalement pour déterminer les variables prédictives. Ensuite, nous comparons les performances des RNA dépendamment de l'algorithme numérique d'apprentissage et le type d'information utilisés pour la prédiction de la tendance future du marché boursier. De même, nous utilisons les machines à support de vecteurs (SVM) pour la prédiction et la comparaison avec les RNA.

En somme et par rapport à la littérature, notre contribution se situe globalement à quatre niveaux. Nous voulons déterminer quelle catégorie d'information classique (économique, technique, historique, comportementale) et quel type d'algorithme d'entraînement du RNA permettent d'obtenir la meilleure performance de prédiction. De plus, nous voulons savoir si la combinaison de toutes ces catégories d'information permettrait d'améliorer la prévision par rapport à chaque type d'information pris séparément. En outre, nous voulons savoir si l'information extraite par la transformée en paquet d'ondelettes de la série temporelle boursière permettrait d'améliorer la prédiction de la tendance future du marché boursier. Finalement, nous comparons les RNA et SVM dans le problème de classification appliqué au domaine de la prédiction boursière.

Noter que la première contribution de cette thèse se situe globalement au niveau de l'examen de l'effet de la catégorie d'information utilisée sur la performance de prédiction du RNA et du SVM. La littérature a ignoré un facteur important dans la détermination de la qualité de la prédiction, soit le type d'information. Or, l'information constitue un intrant considérable en milieux financiers pour la prise de décision d'investissement et de gestion de portefeuille. Dans ce contexte, il est très utile d'examiner l'influence du type d'information sur la qualité de la prévision d'un système prédictif comme le RNA ou le SVM.

Certes, les RNA et SVM sont des techniques qui permettent toutes les deux de classer des données non-linéairement séparables, cependant certaines différences les distinguent. Les premiers ne supposent aucune hypothèse sur la distribution des données, mais ils peuvent tomber dans un minimum local de l'erreur. Par contre, le SVM suppose que les données sont indépendantes et identiquement distribuées, mais il atteint toujours son minimum global. De ce fait, leur comparaison dans le domaine de la prédiction boursière nous paraît importante au niveau pratique; par exemple en milieu professionnel. La littérature a essentiellement comparé les RNA aux modèles statistiques et les conclusions sont bien établies à ce niveau. La comparaison entre RNA et SVM en fonction du type d'information employée pour la prédiction de la tendance future du marché boursier constitue la deuxième contribution majeure de notre thèse.

Le plan de la thèse se présente comme suit. Le premier chapitre passe en revue l'état de l'art. Le deuxième chapitre décrit notre méthodologie. Le troisième chapitre présente les données et les prétraitements. Le quatrième chapitre présente les résultats des simulations. La discussion des résultats obtenus est donnée au niveau du cinquième chapitre. Finalement, nous allons conclure.

Noter que le prétraitement statistique des données est présenté au niveau du troisième chapitre, mais les détails techniques sont présentés en annexe afin d'alléger la lecture du texte.

CHAPITRE I

ÉTAT DE L'ART

Dans ce chapitre, la revue de littérature sur la prévision des prix des indices boursiers est présentée dans la première section. Ensuite, la deuxième section est consacrée à une introduction à la finance comportementale; et en particulier à mettre en évidence la relation entre le sentiment de l'investisseur et le marché boursier.

1.1 Prévision des marchés boursiers

La revue de littérature couvre les applications courantes des techniques intelligentes pour prédire les indices boursiers. En général, les recherches ont utilisé comme variables prédictives l'historique des prix et des volumes, les indicateurs techniques, ou certaines variables macroéconomiques. Toutes ces études ont utilisé ces variables séparément et ont également ignoré l'importance des variables découlant de la psychologie de l'investisseur. En effet, les études qui utilisent les mesures du sentiment de l'investisseur comme variables prédictives ont employé exclusivement les méthodes statistiques afin de déterminer l'existence d'un lien significatif entre le marché boursier et les variables de sentiment, sans pour autant les utiliser pour effectuer des prévisions. Les indicateurs techniques, les variables macroéconomiques et l'historique des rendements passés ont été employés comme entrées à des modèles d'intelligence computationnelles (IC) depuis les années 1990 (Baba et Kozaki, 1992; Brownstone, 1996; Mizuno et al., 1998), mais depuis la décennie 2000, l'intérêt pour la prédiction boursière à l'aide de ceux-ci a connu un grand essor.

Au niveau de la modélisation avec les indicateurs techniques, plusieurs approches d'IC ont été employées, notamment les réseaux de neurones (Lendasse et al., 2000; Motiwalla et Wahab, 2000; Leigh et Purvis, 2002; Bautista, 2001; Dong et al. 2003; Armano et al., 2004; Jaruszewicz et Mandziuk, 2004), le système Neuro-Flou (Baek et Cho, 2002; Tan et al., 2011), la logique floue (Lam, 2001; Dong & Zhou, 2002; Dourra & Siy, 2002), et l'optimisation par essaim et la sélection négative⁸ (Azzini et al., 2010).

Toutes ces recherches ont conclu que les indicateurs techniques offrent un pouvoir prédictif supérieur à la marche aléatoire⁹. Le Tableau 1.1 compare quelques unes des études qui utilisaient les indicateurs techniques comme variables prédictives.

⁸ La sélection négative est une technique qui fait partie des modèles intelligents qui s'inspirent du système immunitaire humain. Voir Azzini et al., (2011) pour plus de détails.

⁹ Random Walk en anglais.

Tableau 1.1

Aperçu sur la littérature : approche par indicateurs techniques

Études	System de prevision	Variables prédictives	Conclusion
Lendasse et al. (2000) BEL20 (Belgique)	RNA à fonction à base radiale Polynomial régression	Rendement passé Différences de rendements passés Indicateurs d'oscillation Moyennes mobiles Moyennes mobiles exponentielles	Taux de classification correct : 65%. Taux de détection des hausses et des baisses sont respectivement 70% et 60%. RNA à fonction à base radiale performe mieux que la régression polynomial.
Baek et Cho (2002) KOSPI (Korea)	Réseau de neurones auto-associatifs	Ratio de volatilité Indice de force relative Différence des moyennes mobiles	Taux de détection des tendances haussières est 95%. Approche proposée performe mieux que la stratégie buy & hold.
Armano et al. (2004) S&P500	RNA localement récurrent Système étendu de classification (système évolutionnaire)	Différences des moyennes Variation Indice de force relative Convexité Tendance haussière Tendance baissière	Système étendu de classification performe mieux que le système RNA localement récurrent : 55.5% taux de classification versus 52.35%.
Azzini et al. (2010) S&P500	Particle Swarm Optimization (PSO) Negative Selection (NS)	Logarithme du rendement Moyenne mobile Moyenne mobile exponentielle Indice de force relative Convergence/divergence Momentum Variation	Le système NS performe mieux que le système PSO : 71% taux de classification versus 67%.
Tan et al. (2011) Singapore & Amérique	Système Neuro-Flou + Apprentissage renforcé	Indice de force relative Moyennes mobiles	Approche proposée permet de battre le marché.

Au niveau des variables macroéconomiques, on peut citer le RNA (Yiwen et al., 2000; Wu et al., 2001; Egeli et al. 2003; Thawornwong et Enke, 2004), et les réseaux Baysiens et les machines à support de vecteurs (Chen et al., 2003; Huang et al., 2005). Les résultats montrent que les variables de type économiques permettent d'obtenir de bonnes performances de prédiction comparativement à la marche aléatoire ou à la régression multiple. Plus récemment, les modèles statistiques ont été également employés pour la prédiction de la tendance future du marché boursier, notamment le S&P500. On peut citer Cengiz et Herwartz (2011) qui ont employé un système prédictif qui hybride la régression partielle et le processus autorégressif, et Nyberg (2011) qui a utilisé le modèle Probit.

Les auteurs ont conclu que les résultats obtenus par ces approches statistiques (régression partielle, processus autorégressif, et modèle Probit) sont encourageants. Le Tableau 1.2 donne un aperçu sur quelques études qui utilisaient les variables économiques comme variables prédictives.

Tableau 1.2

Aperçu sur la littérature : approche par variables économiques

Études & Marché	Système de prevision	Variables prédictives	Conclusion
Wu et al. (2001) S&P500 (Amérique)	Système Neuro-Flou RNA à rétropropagation	Indice de consommation Taux d'intérêt fédéral Indice de production industrielle Taux de chômage Taux d'emprunt bancaire	Sur la base du R carré, le système Neuro-Flou est plus performant dans la prédiction de la tendance future du marché S&P.
Chen et al. (2003) Taiwan	Réseau probabiliste Moindre carré générale + filtre de Kalman Marche aléatoire	Bonds du trésor Taux d'intérêt à court terme Niveau de consommation Produit national brut Produit domestique brut Indice de consommation Indice de production	Le réseau probabiliste performe mieux que les autres techniques et ce pour les différents horizons de prédictions : 3, 6, et 12 mois.
Egeli et al. (2003) Istanbul (Turquie)	RNA à rétropropagation RNA à propagation généralisée Processus à moyenne mobile	Taux de change American Taux d'intérêt Variables dichotomiques pour indiquer le jour de semaine	Les RNA performes mieux que le processus à moyenne mobile. RNA à propagation généralisée performe mieux que le RNA à rétropropagation.
Thawornwong et Enke (2004) S&P500 (Amérique)	RNA à propagation Réseau probabiliste Régression linéaire Marche aléatoire	Bons du trésor Taux d'intérêt fédéral Rendement des obligations fédérales Niveau de production Production industrielle Niveau de consommation Masse monétaire Différence des taux d'intérêt obligataires	RNA à propagation Performe mieux que les autres techniques et aussi que la stratégie de buy & hold.
Huang et al. (2005)	SVM Réseau d'Elman Analyse discriminante Marche aléatoire	Taux d'intérêt à court terme Taux d'intérêt à long terme indice de consommation indice de production produit domestique brut produit national brut consommation publique consommation privée	SVM obtient 73% de taux de classification correcte. Il performe mieux que les autres techniques.

L'historique des rendements boursiers a aussi été utilisé comme entrée aux RNA (Fernandez et al., 2000; Zhang et al., 2004; Yumlu et al., 2005; Constantinou et al., 2006), et aux systèmes à logique floue (Wang, 2002; Atsalakis et Valavanis, 2009, Boyacioglu et Avci, 2010). Les résultats montrent que l'utilisation des rendements passés pour prédire les rendements futurs est effective et permet de surpasser les stratégies standard comme "buy and hold"¹⁰. Une comparaison de quelques études qui utilisaient les indicateurs techniques comme variables prédictives est donnée dans le Tableau 1.3.

¹⁰ Acheter l'actif ou le titre boursier et le garder jusqu'à la fin de la période d'investissement pour le vendre et réaliser un profit.

Tableau 1.3

Aperçu sur la littérature : approche par retards historiques

Études & Marché	Système de prevision	Variables prédictives	Conclusion
Fernandez et al. (2000) Madrid	RNA	9 valeurs passées du prix de clôture	le système performe mieux que la stratégie dite "buy and hold" lorsque le marché est stable ou en phase baissière. Par contre, en phase de tendance haussière du marché boursier la stratégie de "buy and hold" performe mieux que le système utilisé par les auteurs.
Wang (2002) Taiwan	Fuzzy grey	5 valeurs passées du prix de clôture	Système efficace car il requiert peu de données prédictives et que la déviation moyenne des prévisions par rapport aux données réelles est de l'ordre de 9%.
Zhang et al. (2004) Shanghai	RNA	30 valeurs passées du prix de clôture	Taux de classification correcte est 56.3%. Taux de détection des tendances baissières : 85.71%. Le système performe mieux que la stratégie "buy & hold".
Yumulu et al. (2005) Istanbul	RNA multicouches (réseau global) et du RNA récurrent (modèle d'Elman)	7 valeurs passées du prix de clôture	Le réseau global performe mieux que le réseau récurrent selon le taux de classification correct, le taux de détection des tendances haussières, et le taux de détection des tendances baissières.
Atsalakis et Valavanis (2009) Athènes et New York	Système Neuro-Flou	Taux de variations du prix pendant les trois derniers jours	Taux de classification correcte est 68.33 %. le système performe mieux que la stratégie "buy & hold".

Par ailleurs, d'autres études se sont intéressées à l'utilisation des machines à support de vecteurs (Support Vector Machines, SVM) pour améliorer les performances des prévisions. Le SVM introduit par Vapnik (1995) est une méthode à noyau (Kernel) pour l'apprentissage supervisé, à la base fondée sur la théorie statistique. Le SVM est particulièrement indiqué pour la classification binaire. Il exploite une fonction noyau pour effectuer une séparation optimale des données dans un hyperplan. L'hyperplan séparateur optimal est obtenu lorsque la marge qui sépare les données dans leur espace de description est maximisée.

Tai et Cao (2001) ont utilisé le RNA entraîné par descente de gradient et le SVM dans le problème de la prédiction de cinq contrats à termes (Futures) du Chicago Mercantile Market, à savoir les contrats à termes sur le marché S&P500 (CME-SP), les bonds du trésors américain à trente ans (CBOT-US), les bonds du trésors américain à dix ans (CBOT-BO), les bons du trésor du gouvernement allemand à dix ans (EUREX-BUND), et les contrats à termes sur le marché boursier français (MATIF-CAC40). Les entrées pour le SVM et le RNA étaient les retards des prix des contrats à termes à prédire. Les résultats ont montré que la direction future des contrats prédite par le SVM est de l'ordre de 58.29%, 45.22%, 46.73%, 43.21%, et 45.22% respectivement pour CME-SP, CBOT-US, CBOT-BO, EUREX-BUND, et MATIF-CAC40. Par contre le RNA a atteint 55.27%, 44.72%, 41.70%, 41.20%, et 42.22% pour CME-SP, CBOT-US, CBOT-BO, EUREX-BUND, et MATIF-CAC40 respectivement. De ce fait le SVM performait mieux que le RNA.

Dans une autre étude, Cao et Tai (2001) ont comparé les performances du SVM et du RNA entraîné par descente de gradient pour la prévision du prix du S&P500 sur le Chicago Mercantile Market en utilisant de l'information technique comme entrée aux systèmes prédictifs. La direction future du marché est prédite à hauteur de 41.23% par le RNA avec 800 cycles d'apprentissage, à hauteur de 39.20% par le RNA avec 8000 cycles, et à hauteur de 46.23% par le SVM. Les auteurs ont conclu que le SVM

permet de mieux prédire les mouvements du S&P500 que le RNA. De son côté, Kim (2003) a comparé le SVM au RNA et au système de raisonnement à base de cas (RBC) dans la prévision du marché KOSPI en utilisant de l'information technique. Le SVM, le RNA et le RBC ont atteint respectivement 57.83%, 54.73% et 51.97% de taux de prédiction. De ce fait, l'auteur a conclu que le SVM est plus performant que le RNA dans le problème de la prédiction boursière.

Ajith et AuYeung (2003) ont comparé la performance du SVM, système Neuro-Flou (SNF), RNA et du Difference Boosting Neural Network (DBNN) dans la prédiction du NASDAQ et du S&P50 (CNX NiFTY) en utilisant de l'information technique. Pour le NASDAQ, la statistique racine de la moyenne des erreurs au carré (RMEC) obtenue est de l'ordre de 0.0180, 0.0183, 0.0284, et 0.0286 pour le SVM, SNF, RNA, et DBNN respectivement.

Pour le marché S&P50 (CNX NiFTY), la statistique RMEC obtenue par le SVM, SNF, RNA, et DBNN est de l'ordre de 0.0149, 0.0127, 0.0122, et 0.0225 respectivement. De ce fait, les résultats montrent la supériorité du SVM par rapport au RNA entraîné par l'algorithme Levenberg-Marquardt (L-M) et les autres machines employées dans l'étude.

Chen et Shih (2006) ont comparé la performance du SVM et le RNA entraîné avec l'algorithme de la descente du gradient dans la prédiction de six marchés boursiers asiatiques, notamment le Nikkei 225 (NK), All Ordinaries (AU), Hang Seng (HS), Straits Times (ST), Taiwan Weighted (TW) et le KOSPI (KO). Les entrées pour le SVM et le RNA ont été les retards des indices boursiers à prédire. La prédiction du mouvement future du marché obtenue par le SVM est de l'ordre de 57.93%, 65.33%, 54.45%, 55.43%, 55.36%, 54.91% pour le NK, AU, HS, ST, TW, et le KOSPI respectivement. De son côté, le RNA a atteint 59.03%, 56.44%, 61.78%, 54.59%, 52.54%, 74.84% respectivement pour le NK, AU, HS, ST, TW, et le KOSPI. De ce

résultat, les auteurs ont conclu que le SVM est supérieure au RNA avec une exception pour les marchés Nikkei 225 (NK) et Straits Times (ST).

Au niveau de l'analyse multi-résolution des indices boursiers, l'analyse par ondelette a connu un intérêt timide, mais grandissant avec le temps. Son avantage est de décomposer simultanément le signal (série temporelle) dans les domaines temporel et fréquentiel, ce qui permet de faire ressortir deux types d'information : une série d'approximation et une série de détails. Ce type de filtrage est très utile dans le contexte de l'analyse des séries temporelles boursières, qui sont généralement volatiles et non stationnaires (Huang, 2011; Hsieh et al., 2011). De plus, la décomposition de la série temporelle par ondelette permet de distinguer la saisonnalité, d'identifier les ruptures structurelles et les régions de haute volatilité. Or, il est bien connu que les séries temporelles des rendements boursiers sont affectées par ces caractéristiques (Ramsey, 1999; Gençay et al., 2001; Papagiannaki et al., 2005). En effet, l'analyse par ondelette a été employée pour prédire le niveau de l'indice boursier. Par exemple, les transformées (analyse) en ondelettes discrètes ont été utilisées par Aussem et al (1998) et par Murtagh et al (2003) où tous les coefficients sont considérés comme des entrées pour les RNA.

Par contre, Li et al (2006) ont utilisé une transformée en ondelette discrète des séries financières; puis ont calculé neuf¹¹ caractéristiques des coefficients en ondelettes qui sont employées par un système de type programmation génétique pour la prédiction de la tendance future des séries temporelles de l'indice boursier Dow Jones Industrial Average (DJIA): hausse de 2.2% ou baisse de 2.2%. Les auteurs ont conclu que l'analyse par ondelette discrète (AOD) est une approche prometteuse dans la mesure où ils ont réussi à obtenir un taux de prédiction de 62.81% comparativement à

¹¹ Les caractéristiques calculées sont l'énergie, l'entropie, longueur de la courbe, énergie non linéaire, moyenne, minimum, maximum, médiane et écart type.

53.72% atteint lorsque seulement de l'information technique est utilisée comme entrée au réseau récurrent.

Plus récemment, l'analyse par ondelette discrète (AOD) commence à susciter beaucoup d'intérêt pour la prédiction boursière. Par exemple, Huang et Wu (2010) ont employé l'AOD pour extraire les caractéristiques du marché américain, canadien, anglais, français, allemand, espagnol, et japonais. Ensuite, le réseau de type carte auto-organisé est employé pour effectuer un groupement des caractéristiques obtenues pour chaque marché. Finalement, la régression par la méthode des moindres carrée partielle est utilisée pour la prédiction. Les résultats obtenus par les auteurs ont montré qu'en termes de racine de la moyenne des erreurs au carré, l'approche proposée performe mieux que le SVM et le RNA à fonction à base radiale ayant -tout deux- comme entrée les retards des séries à prédire.

Huang (2011) a utilisé l'analyse par ondelette discrète pour décomposer les séries temporelles de plusieurs marchés asiatiques et occidentaux. Ensuite, une régression de type moindre carré partielle est employée pour la prédiction. La performance de cette approche a été comparée à la performance du SVM et du RNA à fonction à base radiale qui, tout deux, utilisent seulement les retards de la série temporelles. L'auteur a conclu que son approche performe mieux par rapport aux SVM et RNA à fonction à base radiale. De leur côté, Huang et al. (2011) ont employé l'AOD pour extraire les coefficients de basse fréquence qui constituent l'entrée au RNA pour prédire le Shanghai Composite Index. Ils ont trouvé que cette approche performe mieux comparativement au RNA dont les entrées sont seulement de l'information historique; par exemple les retards de l'indice boursier.

Hsieh et al. (2011) ont utilisé une transformée en ondelette de type Haar pour obtenir les signaux d'approximation qui sont à leur tour utilisés comme entrées au réseau de neurones récurrent pour prédire le niveau des indices boursiers Dow Jones Industrial Average Index (DJIA), London FTSE-100 Index (FTSE), Tokyo Nikkei-225 Index

(Nikkei), et Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX). Les auteurs ont conclu que leur approche performe mieux que les autres approches proposées dans la littérature pour la prédiction du marché boursier Coréen (KOPSI).

Vue l'intérêt pour l'usage de l'analyse par ondelette discrète ordinaire (AOD) qui s'est manifesté très récemment (Huang et Wu, 2010; Huang, 2011; Huang et al., 2011; Hsieh et al., 2011) à cause de la capacité prédictive des coefficients obtenus par rapport aux information traditionnelles (historique, technique), nous allons simulé le même modèle de Huang et al. (2011) et allons comparer sa performance par rapport au RNA qui utilise de l'information extraite par la nouvelle génération d'ondelettes; notamment l'analyse multi-résolution par paquets d'ondelettes (APO). Cette dernière permet de mieux analyser et caractériser la série temporelle en question en décomposant à la fois ses basses fréquences et ses hautes fréquences.

Par ailleurs, la littérature a également montré l'utilité des machines à support de vecteurs (SVM¹²) dans la prédiction de la tendance du marché boursier. En effet, Tai et Cao (2001), Ajith et AuYeung (2003), et Kim (2003) ont conclu que le SVM est plus performant que le RNA dans le problème de la prédiction boursière. Par contre, Chen et Shih (2006) ont trouvé que dépendamment du marché à prévoir, le SVM peut ou ne pas être supérieure au RNA. Par ailleurs, Lai et al. (2010) et Kara et al. (2011) ont conclu que le RNA performe mieux que le SVM. De ce fait, nous voulons vérifier la supériorité du SVM par rapport au RNA. En fait, il s'agit d'un objectif secondaire de notre thèse.

¹² Les machines à support de vecteurs vont être employées au singulier en utilisant le terme SVM pour alléger et simplifier le texte.

1.2 Finance comportementale et sentiment de l'investisseur

Les théories normatives qui utilisent la fonction d'utilité tiennent compte du profil de l'individu (investisseur) et de son patrimoine. Par exemple, le Capital Asset Pricing Model (CAPM¹³) (Sharpe, 1964; Lintner, 1965; Mossin, 1966) est l'un des résultats centraux de la théorie financière moderne. Il est principalement basé sur la rationalité de l'investisseur. La notion de rationalité renferme deux hypothèses quant au comportement de l'individu. La première suppose une révision des croyances en cas d'information nouvelle conforme à la règle de Bayes. La deuxième stipule la stabilité des préférences et la maximisation de l'utilité espérée.

Cependant, le profil psychologique de l'investisseur ou du marché boursier est ignoré, car l'hypothèse fondamentale est que son processus de prise de décision est rationnel et est cohérent. Donc, les théories normatives du processus de prise de décision ne sont pas toujours adaptées au fonctionnement de l'esprit humain.

Les études empiriques en finance (Herbert, 1987; Barberis et al., 1998; Kahneman and Riepe; 1998, Shapira and Venezia, 2000; Hirshleifer, 2001; Shefrin, 2002) ont montré l'existence de certaines "anomalies" : comportement irrationnel de l'investisseur/individu. Pour expliquer ces anomalies, Kahneman et Tversky (1979) ont proposé la finance comportementale (Behavioral Finance) comme une alternative. La finance comportementale repose sur deux piliers indissociables que sont la psychologie de l'investisseur d'une part, et les limites de l'arbitrage d'autre part.

La thèse selon laquelle l'investisseur individuel n'est pas entièrement rationnel a été confirmée par de nombreuses études expérimentales (Kahneman et Tversky, 1979 ; Barberis et al., 1998 ; Daniel et al., 1997). Selon (Daniel et al., 2001; Hirshleifer, 2001; Odean, 1998), les investisseurs individuels dévient systématiquement du comportement optimal; ce qui peut affecter l'évaluation des actifs et donc les

¹³ Voir annexe I aussi appelé le modèle d'évaluation des actifs.

prévisions. Plusieurs études attribuent ces “biais” (ou anomalies) à l’excès de confiance¹⁴ (Barber et Odean, 2000; Gervais et Odean, 2001), la sur-réaction¹⁵ (DeBond et Thaler, 1985), l’aversion à la perte (Odean, 1998; Shefrin et Statman, 1985; Kahneman et Tversky, 1979), et la mauvaise calibration des probabilités (Lichtenstein et al., 1982), etc.

D’un autre côté, beaucoup d’études ont examiné comment les intervenants du marché de type “noisy traders” (spéculateurs) exploitent des signaux bruités comme le sentiment ce qui affecte les prix des actifs. En effet, De Long et al., (1990) ont montré que si les arbitragistes averses au risque savent comment les prix divergent de leurs valeurs fondamentales avant d’y converger à nouveau, ils peuvent prendre des petites positions pour parier contre les prix en déséquilibre. Ainsi, si ce type d’intervenants décide de leurs transactions sur la base du sentiment, alors les mesures du sentiment peuvent avoir un pouvoir prédictif du comportement des prix des actifs. Et donc, le sentiment peut prédire les rendements et/ou la volatilité des actifs financiers transigés dans les bourses des valeurs.

En effet, la tendance actuelle en finance comportementale et en psychologie du marché se concentre davantage sur l’étude du sentiment dans la prise de décision (Ben-Rephael et al., 2010; Chi et al., 2011; Hribar et McInnis, 2011; Yang et Wu, 2011).

Mesuré par l’enquête de l’association des investisseurs, De Bondt (1993) montre que le sentiment permet de prévoir le rendement futur avec fiabilité. Neal et Wheatley (1998) utilisent la décote sur les fonds à capital fermé et le rachat net des fonds mutuels comme étant des mesures du sentiment. Ils ont découvert que ces mesures

¹⁴ Les investisseurs ont tendance à avoir de plus en plus confiance à leurs anticipations sur la base de leurs performances antérieures ou l’information dont ils disposent.

¹⁵ Les investisseurs réagissent excessivement à de nouvelles informations.

peuvent aider à prévoir le rendement des compagnies à petites capitalisations. Mesuré par le l'indice de volatilité VIX (Volatility Index) du Chicago Board Options Exchange (CBOE), Whaley (2000) a conclu que le sentiment affecte significativement les indices boursiers S&P100 et le Dow Jones Industrial Average. Fisher et Statman (2000, 2003) ont trouvé que les mesures du sentiment chez les investisseurs individuels et les professionnels de l'industrie de la finance (comme l'indice American Association of Individual Investors, AAI) permettent de prévoir les rendements du S&P500. Simon et Wiggins (2001) ont utilisé un indice de la volatilité implicite, le ratio Put-Call et l'indice Arms comme mesures du sentiment. Les résultats empiriques montrent que ces mesures aident à prévoir la rentabilité du marché boursier. Par ailleurs, Warther (1995), Edwards et Zhang (1998), et Fant (1999) ont utilisé l'indice ICI (Investor Confidence Index (ICI)) qui reflète le flux et la valeur moyenne des fonds mutuels américains comme mesure de sentiment. Les résultats ont montré que le sentiment affecte la tendance du marché boursier. D'autres études ont également confirmé la capacité prédictive des mesures du sentiment (Wang, 2001, 2003; Charoenrook, 2003; Lemmon et Portniaguina, 2004; Qiu et Welch, 2004; Baker and Wurgler, 2000, 2004).

Plus récemment, Bandopadhyaya et Jones (2008) ont trouvé que le ratio put/call et l'indice de volatilité VIX affectent largement le marché S&P500. Ben-Rephael et al. (2010) ont de leur côté utilisé le VIX et le ICI comme mesures de sentiment. Les résultats ont montré que le sentiment affecte trois marchés américains, notamment le NYSE, Amex, et Nasdaq. Chi et al., (2011) ont utilisé le flux des fonds mutuels comme mesure de sentiment et ont trouvé que ce dernier affecte significativement le marché boursier chinois. Hribar et McNinnis (2011) ont conclu que le sentiment (mesuré par la moyenne de la clôture des fonds mutuels, le turnover des actions du marché boursier NYSE et les offres publiques d'achat) est positivement associé aux prédictions des rendements.

Yang et Wu (2011) ont utilisé le ratio put/call, le volume de transactions et le rapport des ordres des transactions (achat/vente) comme mesures de sentiment. Ils ont trouvé que ces mesures du sentiment affectent significativement la volatilité du marché taïwanais. Ce résultat a été déjà constaté par Barber et al. (2009), Yang et Wu (2011), et Yu et Yuan, (2011) qui ont montré que le sentiment est fortement lié à la volatilité du marché boursier.

En effet, le risque est communément considéré comme un hasard (March et Shapira, 1992); par exemple la probabilité de perte, revenu négatif ou en dessous d'une limite. La théorie financière moderne définit le risque comme étant une variation du rendement. De ce fait, la volatilité est définitivement la mesure de risque acceptée et répandue dans la communauté académique et praticienne de la finance (Lawarence et Makridakis, 1989 ; Yu et Yuan, 2011). En outre, les séries temporelles de la volatilité des prix historiques peuvent avoir un impact significatif sur le comportement des prévisions de l'investisseur (Lawarence et Makridakis, 1989 ; Yu et Yuan, 2011). En effet, mesuré par la volatilité du marché boursier, Wang et al., (2006), Schmeling (2007) et Verma et Verma (2007) ont conclu que le sentiment affecte les anticipations des investisseurs.

Cependant, les résultats de la littérature font ressortir deux conclusions différentes concernant le rôle de la volatilité dans la prévision des rendements futurs. Par exemple, les études empiriques ont montré l'existence d'une relation positive entre la volatilité et le rendement (Goyal et Santa Clara, 2003 ; Malkiel et Xu, 2003) ou d'une relation négative (Ang et al., 2005).

En somme, il faudrait noter que le sentiment affecte le comportement de prise de décision chez l'investisseur selon deux avenues. La première est que le sentiment joue un rôle dans l'évaluation de l'actif (Brown et Cliff, 2004, 2005; Baker et Wurgler, 2007; Verma et al., 2008; Ho et Hung, 2009). La deuxième est que le sentiment est associée à la volatilité du titre (Barber et al., 2009).

1.3 Le réseau de neurones artificiels spatio-temporel

La prédiction spatio-temporelle gagne de plus en plus de l'importance grâce à sa performance prometteuse dans le traitement de données complexes dans lesquelles les caractéristiques qui doivent être prises en considération sont spatiales et temporelles. La prédiction spatio-temporelle et la technique d'analyse de tendance sont des moyens efficaces pour prédire l'attribut spatial. La prédiction spatio-temporelle a été développée conjointement à partir de deux domaines d'études : la prédiction spatiale et la prédiction temporelle. En particulier, les séries temporelles sont projetées dans un domaine spatial et les données spatiales sont projetées dans un domaine temporel. En d'autres mots, il s'agit de la combinaison de l'information spatiale et de l'information temporelle pour modéliser et prédire une série temporelle donnée. Le défi de prédiction spatiotemporelle est de savoir comment intégrer l'espace et le temps de façon homogène et simultanée. Plusieurs études ont tenté de combiner l'information temporelle (séries temporelles) et l'information spatiale en utilisant des modèles de régression linéaires (Deutsch et Ramos 1986, Pfeifer et Deutsch 1990, Cressie et Majure 1997, Pokrajac et Obradovic 2001, Cheng et Wang 2006, Cheng et Wang 2007). Cependant, l'utilisation d'une telle régression linéaire est juste une simplification d'associations spatio-temporelles compliquées qui existent dans les phénomènes physiques complexes. De ce fait, les réseaux de neurones spatio-temporels (Unnikrishnan, Hopfield, et Tank, 1991; Chung, Chen, et Wu, 1998) ont été développés pour traiter les informations temporelles et spatiales d'un processus non stationnaire. Les réseaux de neurones spatio-temporels ont été appliqués avec succès dans des applications où le temps et l'espace sont dominants. On peut citer par exemple la modélisation de système dynamique de type MIMO (Multiple In Multiple Out) (Atiya, 1992), la reconnaissance d'objet de scène (Chung, Chen, et Wu, 1998), l'imagerie médicale (Chang, Chung, et Lai, 2002), la prédiction du trafic (Hu, Xie, Song, Wu, 2008), le contrôle de stimuli en interface homme-machine bi-directionnelle (Liu, Khalil, et Oweiss, 2011), et plus récemment la transmission sans

fil (Rawat et Ghannouchi, 2012). Cependant, l'application des réseaux de neurones spatio-temporels en prédiction boursière est rare ou inexistante. Ceci peut être expliqué par le fait que, à l'instar de notre sujet de thèse, la littérature en prédiction financière se consacre uniquement à la modélisation des séries temporelles univariées qui ne disposent pas d'information spatiales; par exemple l'indice boursier S&P500 considéré dans notre thèse est géographiquement localisé dans un seul marché. De ce fait, la dimension ou l'information spatiale n'existe pas dans ce genre d'étude. Néanmoins, le réseau de neurones spatio-temporel pourrait être appliqué dans des cas où les informations temporelles et spatiales peuvent caractériser les données boursières. Plus précisément, pour nos travaux futurs, nous prévoyons de modéliser un titre boursier côté simultanément dans plusieurs places boursières mondiales. Ou encore, modéliser l'interaction de plusieurs marchés boursiers internationaux comme un seul système dynamique sachant que la littérature en finance empirique a montré que les marchés boursiers internationaux sont co-intégrés (Ellul, 2006; Cipollini et Kapetanios, 2009, Conrad, Karanasos, et Zeng, 2011; Turtle et Zhang, 2012). Dans ce dernier cas, le réseau de neurones spatio-temporel permettra de modéliser les interactions entre les différents marchés boursiers et donc de mieux prédire leurs mouvements futurs. Il s'agira de la modélisation spatio-temporelle d'un système dynamique. Ce travail fera partie de nos travaux futurs et non pas dans le cadre de l'actuelle thèse pour les raisons qui vont être présentées au niveau de la contribution de la recherche.

1.4 Le chaos dans la prédiction boursière

Un système chaotique est un système dynamique qui possède une forme non-linéaire forte pouvant engendrer des fluctuations non périodiques, des cycles mixtes et des variations dans les données (Guégan, 2009). En particulier, ses états d'équilibre peuvent être des orbites qui peuvent servir à faire des prévisions structurées. Pour cette raison, des efforts ont été menés pour l'utiliser en finance en vue d'obtenir des prévisions plus précises (Chen, 2006).

Dans le problème de la prédiction des séries temporelles financières et pour des raisons pratiques, le chaos est supposé être un processus non-linéaire composé d'une partie déterministe et d'une autre partie aléatoire (Baker, 1996; Alligood et al., 1997; Devaney, 2003; Sprott, 2003). Le processus déterministe peut être caractérisé en utilisant une régression, tandis que le processus aléatoire peut être caractérisé par les paramètres statistiques d'une fonction de distribution. Ainsi, en utilisant seulement les processus déterministes ou aléatoires on ne pourra pas capturer entièrement la nature d'un système chaotique.

Dans le cadre de la prédiction boursière, la théorie du chaos est une approche relativement nouvelle pour la modélisation de systèmes dynamiques non-linéaires comme le marché boursier. L'analyse de chaos a été utilisée pour analyser la dynamique non-linéaire dans la série temporelle (Embrechts, Cader, et Deboeck, 1994; Garliauskas, 1999; Gang, Zhongke, et Guanzhong, 2000; Oh et Kim, 2002).

D'autres études ont rapporté que les RNA fournissent de meilleurs résultats comparés aux approches statistiques quand la série temporelle boursière est chaotique (Wang et Leu, 1996; Kamruzzaman et Sarker, 2003; Moeini, Ahrari, et Karimi, 2012); car selon les auteurs, le réseau de neurone peut capturer la composante déterministe et la composante aléatoire. Cependant, la prédiction à long terme des systèmes chaotiques est presque impossible parce qu'une petite erreur peut s'amplifier au fil du temps pour causer de nouveaux événements et donc des divergences (Garliauskas, 1999; Gang, Zhongke, et Guanzhong, 2000; Moeini, Ahrari, et Karimi, 2012). En effet, avec la présence de chaos les prédictions à long terme ne sont pas faisables car un léger écart dans les valeurs des principaux paramètres du modèle chaotique aboutirait à une large déviation dans les valeurs des séries temporelles ce qui augmentera les erreurs (Garliauskas, 1999; Gang, Zhongke, et Guanzhong, 2000; Moeini, Ahrari, et Karimi, 2012). Autrement dit, le système chaotique est extrêmement sensible aux conditions initiales et les erreurs s'accumulent vite ce qui rend en pratique la prédiction à long

terme non possible. Dans ce cas-ci, seulement la prédiction à court terme est possible (Garliauskas, 1999; Gang, Zhongke, et Guanzhong, 2000; Moeini, Ahrari, et Karimi, 2012). Dans cette thèse, la série temporelle des rendements boursiers du S&P500 qui fait l'objet de notre étude ne montre pas de comportement chaotique évident (voire la Figure 11). Par contre, la série semble clairement suivre un processus stochastique dont la moyenne conditionnelle est asymptotiquement constante et la variance conditionnelle asymptotique dépend du temps. De ce fait, une modélisation chaotique ne semble pas adaptée à nos données.

D'un point de vue pratique, dans le cadre d'une gestion de portefeuille passive, un gestionnaire de portefeuille a besoin de faire des prédictions à long terme des titres boursiers pour pouvoir faire une allocation d'actifs stratégique (optimisation de portefeuille) plus appropriée. Par contre, dans le cas où le gestionnaire suit une stratégie de portefeuille active et donc à court terme, alors une analyse du comportement chaotique potentiel des séries temporelles pourrait se révéler bénéfique.

Nous envisageons procéder à une telle analyse dans l'avenir pour tester sa capacité de prévision à court terme dans le cadre de la gestion d'un portefeuille simulé. Pour l'heure, notre recherche veut d'abord examiner l'effet du type d'information sur le réseau de neurones dans le cadre de la prédiction de la tendance future du marché.

1.5 la sélection des variables prédictives

La sélection des caractéristiques ou bien des variables prédictives est une étape importante dans les systèmes de classification dans la mesure où elle permet de réduire le nombre de caractéristiques tout en essayant de préserver ou d'améliorer la performance du classifieur utilisé. Il s'agit de choisir un sous-ensemble minimum de caractéristiques à partir d'un ensemble original de sorte à ce que l'espace de caractéristiques soit réduit de façon optimale selon certains critères d'évaluation (Liu et Yu, 2005). Puisque notre thèse est concernée par la sélection des caractéristiques

(variables) pertinentes à la modélisation et prédiction des tendances futures du marché boursier, alors il est important de présenter un survol sur les techniques de sélection des caractéristiques couramment utilisées dans le domaine de classification et reconnaissance de formes. En effet, il est bien connu que la performance d'un algorithme d'apprentissage automatique supervisé en termes d'exactitude et de temps de calcul se dégrade quand beaucoup de caractéristiques inutiles sont utilisées comme entrées aux systèmes prédictifs (Cost et Salzberg, 1996; John et al., 1994). Alors, une question fondamentale dans l'apprentissage automatique est comment choisir un sous-ensemble approprié de caractéristiques qui permettrait d'améliorer la performance du classifieur. En effet, quand la base de données est grande et contient beaucoup de caractéristiques alors il est fortement recommandé d'éliminer celles qui pourraient introduire du bruit au système de classification du fait qu'elles ne sont pas statistiquement reliées à la variable à prédire. Ceci peut être dû au fait qu'elles sont sans relation avec la variable à modéliser ou bien qu'elles sont corrélées entre elles. En somme, la détermination des entrées pertinentes au système de classification est non seulement importante pour améliorer la performance du classifieur mais également pour être en mesure de comprendre les données et d'interpréter les résultats (Kohavi and Sommerfield, 1995). De plus, le temps d'exécution de l'algorithme d'apprentissage peut considérablement être diminué (Koller et Sahami, 1996).

Le choix des variables prédictives pertinentes peut être réalisé par des experts humains qui sont spécialistes du domaine d'application en question; autrement le problème de haute dimension de l'ensemble qui contient les caractéristiques (entrées) devient difficile à résoudre. Cependant, plusieurs techniques automatiques ont été développées dans la littérature dans le domaine de la sélection de caractéristiques. Ce domaine est devenu un champ de recherche très actif depuis quelques années.

Les méthodes de sélection des caractéristiques peuvent être séparées en trois types de catégories (Saeys et al. ,2007): les approches de type filtre, de type symbiose (Wrapper), et de type embarqué (Embedded). Elles sont respectivement présentées dans les Tableaux 1.4, 1.5, et 1.6 qui sont adaptés de Saeys et al. (2007). Les méthodes de l'approche filtre exploitent les propriétés intrinsèques des caractéristiques utilisées et elles sont indépendantes du classifieur. Elles permettent aussi de modéliser les dépendances entre les caractéristiques. Les méthodes de type symbiose déterminent la pertinence des caractéristiques selon la performance du classifieur. De ce fait, elles sont plus coûteuses en termes computationnels que les approches de type filtre; mais elles sont plus efficaces. Finalement, les approches de type embarqué (Embedded) sont spécifiques pour un algorithme d'apprentissage donné et elles ont l'avantage d'interagir avec le modèle de classification et d'être moins coûteuses que les approches de types symbiose.

Tableau 1.4

Techniques de sélection : approche de type filtre

Avantages	Inconvénients	Exemples
<u>Univarié</u> Rapide Évolutif Indépendant du classifieur	Ignore les dépendances entre les caractéristiques Ignore les interactions avec le classifieur	χ^2 Distance euclidienne t-test. Gain d'information
<u>Multivarié</u> Modélise les dépendances entre les caractéristiques Indépendant du classifieur Complexité computationnelle réduite par rapport aux méthodes de type Wrapper	Lent par rapport aux techniques univariées Moins évolutif que les techniques univariées Ignore les interactions avec le classifieur	Sélection à base de corrélation Markov blanket filter Sélection rapide à base de corrélation

Tableau 1.5

Techniques de sélection : approche de type Wrapper

Avantages	Inconvénients	Exemples
<u>Déterministique</u>		
Simple	Risque de sur-apprentissage	Sequential forward selection
Interagit avec le classifieur		
Modélise les dépendances entre les caractéristiques	Risque d'atteindre un minimum local	Sequential backward elimination
Complexité computationnelle réduite par rapport aux méthodes aléatoires	Sélection dépend du classifieur	Beam search
<u>Aléatoire</u>		
Modélise les dépendances entre les caractéristiques	Risque d'atteindre un minimum local	Simulated annealing
Indépendant du classifieur		Randomized hill climbing
	Risque de sur-apprentissage	Algorithmes génétiques
	Coût computationnel élevé	

Tableau 1.6

Techniques de sélection : approche de type embarqué (embedded)

Avantages	Inconvénients	Exemples
Interagit avec le classifieur	Sélection dépend du classifieur	Arbres de décisions
Modélise les dépendances entre les caractéristiques		Bayes naïf pondéré
Complexité computationnelle réduite par rapport aux méthodes de types wrapper		

La plupart des études précédentes qui sont mentionnées dans la littérature et qui utilisent les réseaux de neurones artificiels pour la prédiction des prix des indices des marchés boursier n'ont pas utilisé une technique pour la sélection des variables prédictives pertinentes. Cependant, quelques unes ont utilisé les critères d'information Akaike et le critère Baysien d'information pour la sélection des retards (Qi & Zhang, 2001), l'analyse en composante principale (Ajith et al., 2003), et l'approche wrapper en utilisant les algorithmes génétiques (Huang et al, 2012).

En principe, pour comprendre la direction de flux de l'information dans un système de classification il y a de l'importance fondamentale pour distinguer le conducteur (variables prédictives) du destinataire (variable à prédire). Le test de causalité proposé par Granger (1969) est probablement la méthode la plus en vue d'évaluer la direction d'influence causale dans le temps dans le cadre de l'analyse des séries temporelles. Le test de causalité de Granger permet d'examiner les interactions entre deux variables uni-variées; et en particulier déterminer le sens de la relation causale entre les variables en question. La base fondamentale des évaluations des relations causales utilisant le test de causalité de Granger est le fait qu'une cause (variables

prédictives) précède l'effet (comportement de la variable à prédire). Par conséquent, le test de causalité proposé par Granger (1969) serait plus utile pour la sélection des variables prédictives malgré le fait qu'il est basé sur des régressions linéaires bivariées. Nous allons expliquer pourquoi il est retenu pour la sélection des variables prédictives au niveau de la section qui traite de la contribution de la recherche.

1.6 Critique et proposition de la recherche

Le problème de la prédiction des rendements boursiers est largement documenté dans la littérature. Cependant, certaines critiques devraient être mentionnées comme suit :

La première critique porte sur le fait qu'aucune étude ne s'est penchée sur une comparaison directe des performances des différentes catégories d'informations; en particulier les informations de type économique, technique, ou historique qui sont les plus utilisées comme entrées aux outils de prédiction en général; et en particulier pour les réseaux de neurones. En effet, chaque étude prétend avoir de meilleurs résultats mais on ne connaît pas son efficacité par rapport aux autres catégories d'information.

La deuxième critique concerne le fait que les études antérieures qui ont utilisé les réseaux de neurones n'ont pas considéré les variables prédictives de type psychologique en l'occurrence les mesures de sentiment de l'investisseur qui sont bien documentées dans le domaine de la psychologie de l'investisseur. En effet, la littérature de la finance comportementale utilise seulement les modèles statistiques linéaires de type régression de moindre carré ordinaire pour examiner la relation entre le sentiment de l'investisseur et le marché boursier (Wang, 2001, 2003; Charoenrook, 2003; Lemmon et Portniaguina, 2004; Qiu et Welch, 2004; Baker and Wurgler, 2000, 2004; Brown et Cliff, 2004, 2005; Baker et Wurgler, 2007; Verma et al., 2008; Ho et Hung, 2009; Barber et al., 2009; Ben-Rephael et al., 2010; Chi et al., 2011; Hribar et McInnis, 2011; Yang et Wu, 2011). Or, les méthodes statistiques linéaires sont bien limitées par les hypothèses de la linéarité et de la normalité des distributions. De plus, toutes les études du domaine de la psychologie de l'investisseur ne considèrent pas un

échantillon de test afin d'effectuer des prévisions qui permettent de valider le pouvoir prédictif des mesures de sentiment. En somme, les mesures de sentiment de l'investisseur pourraient constituer un ensemble d'entrées bien utile pour la modélisation et la prévision des séries temporelles boursières par le réseau de neurones artificiels.

La troisième critique concerne le non usage d'un ensemble qui agrège toute l'information disponible à l'investisseur. En particulier, l'effet de la combinaison de toutes les catégories d'information (économique, technique, retard, sentiment) sur le pouvoir prédictif du réseau de neurones demeure inconnu. En d'autres mots, il se pourrait que si l'on considère toutes les catégories d'information dans la modélisation et la prédiction de la tendance future du marché boursier; alors on pourrait obtenir une meilleure performance de prévision.

La quatrième critique se situe au niveau du choix de l'algorithme numérique pour l'apprentissage du réseau de neurones (RNA) par descente du gradient. En effet, l'algorithme de Levenberg-Marquardt (L-M) est le plus utilisé dans l'entraînement des RNA dans le domaine de la prédiction boursière (Atsalakis et Valavanis, 2009; Wang et al., 2011). Cependant, il existe d'autres algorithmes numériques pour la résolution des problèmes de minimisations non linéaires et qui n'ont pas été considérés dans la littérature.

La cinquième critique se situe au niveau du choix des entrées pour les systèmes prédictifs de type réseau de neurones. En effet, dans la revue de littérature la sélection pertinente des entrées est rarement employée. En général, un large ensemble de variables prédictives de différentes catégories est considéré sans pour autant vérifier si ces variables d'entrées causent une variation dans les rendements. Cette approche risque d'inclure automatiquement des entrées qui introduisent du bruit dans les systèmes prédictifs et donc réduisent leurs performances de prédictions. Dans le

meilleur cas, l'analyse par composantes principales est utilisée pour réduire la taille de l'ensemble des entrées (Ajith et al., 2003).

La sixième critique concerne l'usage de l'analyse (transformée) en ondelettes ordinaire pour le traitement des séries temporelles boursières. Malgré que les auteurs des études précédentes (Aussem et al., 1998; Ramsey, 1999; Gençay et al., 2001; Murtagh et al., 2003; Papagiannaki et al., 2005; Li et al., 2006; Huang et Wu, 2010; Huang, 2011; Huang et al., 2011; Hsieh et al., 2011) ont bien souligné le rôle effectif de l'analyse par ondelette dans le traitement des séries boursières pour effectuer les prévisions, il faudrait noter que la transformé en ondelette ordinaire est sous-optimale car elle n'effectue pas une analyse multi-résolution dans toutes les sous-bandes possibles. En effet, une meilleure information à extraire peut exister dans d'autres sous-bandes fréquentielles. Pour obtenir une meilleur analyse multi-résolution et donc extraire de l'information pertinente, il serait plus avantageux d'utiliser la nouvelle génération d'ondelette appelée analyse en paquet d'ondelette. Cette dernière effectue une décomposition multiple des sous-bandes fréquentielles sous une contrainte d'optimisation.

Les critiques précédentes nous amènent à formuler les questions suivantes auxquelles nous voudrions répondre dans le cadre de notre thèse :

Question.1 :

Quel type d'information pourrait fournir les meilleures prédictions ?

Question.2 :

Est ce qu'une combinaison des différents types d'information permettrait d'améliorer les prévisions ?

Question.3 :

Quel algorithme d'approximation numérique permet un meilleur apprentissage au RNA et donc de mieux modéliser la série temporelle et fournir les meilleures prédictions ?

Question 4 :

Est ce qu'on peut améliorer les résultats par rapport aux catégories d'information traditionnelles en considérant de l'information temps-fréquence obtenue par l'analyse multi-résolution; en l'occurrence analyse par paquets d'ondelettes ?

1.7 Contribution de la recherche

La contribution de ce travail se situe à cinq niveaux :

Premièrement, on voudrait vérifier quelle catégorie d'information (économique, technique, comportementale, historique) permettrait de minimiser l'erreur de prévision du RNA, et si ces quatre catégories d'information pourraient conjointement minimiser davantage l'erreur de prévision du RNA. Ce dernier est choisi car il est le plus utilisé dans les applications de l'intelligence artificielle en finance (Mochón et al., 2008; Atsalakis et Valavanis, 2009; Wang et al., (2011); Guresen et al., 2011).

Deuxièmement, nous voulons examiner l'impact du choix de l'algorithme numérique d'apprentissage sur la performance de prédiction du RNA.

Troisièmement, nous allons utiliser des mesures de sentiment des investisseurs comme entrées aux systèmes prédictifs (RNA et SVM). En effet, la littérature a montré que les mesures de sentiment des investisseurs reflètent l'état psychologique du marché boursier et montre aussi qu'elles permettent de prédire le marché en utilisant la régression linéaire (Wang, 2001, 2003; Charoenrook, 2003; Lemmon et Portniaguina, 2004; Qiu et Welch, 2004; Baker and Wurgler, 2000, 2004; Brown et

Cliff, 2004, 2005; Baker et Wurgler, 2007; Verma et al., 2008; Ho et Hung, 2009; Barber et al., 2009; Ben-Rephael et al., 2010; Chi et al., 2011; Hribar et McNinnis, 2011; Yang et Wu, 2011).

Quatrièmement, nous allons utiliser l'analyse en paquet d'ondelette (APO) (Coifman et Wickerhauser, 1990; Eshaghi et Mollaei, 2010; Zhao et Ye, 2010) pour obtenir une meilleure décomposition de la série temporelle boursière. Les signaux d'approximation et de détails constituent les entrées au RNA. Nous appelons ce modèle APO-RNA. L'objectif est de savoir si l'information fréquentielle permettrait d'obtenir une meilleure prédiction que l'information de type économique, technique, historique (retards), ou mesures de sentiments de l'investisseur.

Cinquièmement, nous allons comparer les performances du SVM et du RNA dépendamment (i) de l'algorithme numérique employé pour l'entraînement du RNA et (ii) de la catégorie d'information utilisée comme entrée au RNA et au SVM.

En somme, la contribution de ce travail au niveau informatique repose sur l'examen de la performance des algorithmes numériques d'apprentissage des RNA en termes de modélisation des séries temporelles boursières. En effet, la littérature est riche en matière d'utilisation du RNA en prédiction des indices boursiers, mais l'algorithme le plus utilisé (Dong et Zhou, 2008; Danial, 2009; Jha et al., 2009; He et al., 2009; Eleftherios, 2010; Golovachev, 2011; Li, 2011) est ce lui de la descente du gradient selon la méthode de Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1966; Nocedal et Wright, 2006; Eligius et Boglárka, 2010). Or, il existe plusieurs d'autres algorithmes numériques dont il faudrait examiner l'efficacité en prédiction boursière; comme le Quasi-Newton (Nocedal et Wright, 2006; Eligius et Boglárka, 2010) de Broyden (1970), Fletcher (1970), Shanno (1970) et Goldfarb (1970) (BFGD), l'algorithme du gradient conjugué de type Powell-Beale (Powell, 1977; Eligius et Boglárka, 2010), Fletcher-Reeves (Scales, 1989; Nocedal et Wright, 2006) ou Polak-Ribière (Scales, 1985; Nocedal et Wright, 2006).

Au niveau cognitif, nous utilisons comme entrées aux RNA des variables quantitatives qui renseignent sur l'État psychologique du marché tel que proposé par la finance comportementale. De plus, nous effectuons un choix judicieux des variables prédictives qui sont disponibles à l'investisseur. En d'autres mots, nous utilisons une méthodologie basée sur le test de Granger (1969) qui est simple et efficace pour effectuer le choix des variables prédictives pertinentes par l'investisseur parmi un ensemble d'information qui lui est disponible.

En effet, l'utilisation du test de Granger comme approche pour sélectionner les variables prédictives offre six avantages majeurs : (i) Il permet d'éviter de sélectionner les variables prédictives qui sont empiriquement influencées par la variable à prédire. En effet, il est important d'éviter ce genre de "feedback" car les résultats peuvent diverger dans un système de prédiction dont les variables prédictives ont une relation réciproque avec la variable à prédire (Granger, 1969). (ii) Le fait d'éliminer tout feedback revient à régler le problème d'endogénéité (interactions entre les variables) des données ce qui permet par conséquent de réduire (voir éliminer) le bruit causé l'influence de la variable à prédire sur les variables prédictives. En effet, ce genre de bruit causé par le "feedback" de la variable à prédire risque d'affecter négativement la performance du classifieur. (iii) De fait, le test de Granger permet de composer un ensemble cohérent de données prédictives (Kayse, Sun, D'Esposito, 2009). (iv) Le test de Granger est basée sur de l'inférence statistique et donc permet de sélectionner les variables prédictives avec une probabilité d'occurrence prédéterminée. (v) Le dernier avantage réside dans le fait que le test de Granger permet d'identifier les variables prédictives qui sont en conformité avec la théorie économique et financière (Harvey, 2001; Marquering et Verbeek, 2004; Ludvigson et Ng, 2007; Engle et Rangel, 2008; Campbell et Diebold, 2009; Lettau et Ludvigson, 2010; Avramov et al., 2011). Finalement, (vi) le test de causalité de Granger est devenu un outil standard en économétrie appliquée en

générale et en économétrie financière en particulier (Kumar et Lee, 2006; Tetlock, 2007; Hartwig, 2010; Lux, 2010; Jinjarak et Sheffrin, 2011; Lux, 2012).

Certes, le test de causalité de Granger est basé sur une régression linéaire bi-variée alors que les données traitées sont non-linéaires en général. Pour le rendre statistiquement valide, nous allons prétraiter les données pour qu'elles soient stationnaires et linéaires en moyenne. De ce fait, les résultats du test seront statistiquement valides à un niveau de signification prédéterminé (exemple; la bonne variables prédictive est identifiée à 95% de certitude). De plus, nous faisons l'hypothèse implicite que les conclusions tirées du test de Granger sont aussi valides au niveau non-linéaire. Malheureusement, il n'existe pas de test statistique non-linéaire¹⁶ qui puisse déterminer une variable prédictive avec certitude (par inférence probabiliste).

En somme, l'objectif est de déterminer les variables (économiques, techniques, ou de sentiment) qui statistiquement influencent le comportement de l'indice boursier. En d'autres mots, pour chaque catégorie d'information standard (économique, technique, mesures de sentiment des investisseurs) seules les variables qui montrent un lien causal avec le marché boursier seront retenues.

En effet, le principal objectif de notre thèse est de savoir d'une part quelle catégorie d'information est la plus utile parmi toute l'information publiquement disponible et

¹⁶ Nous avons également testé une approche non statistique, et "non-linéaire", pour la sélection des variables basée sur l'examen des poids du réseau de neurones artificiels : les variables dont le poids est grand (par exemple supérieure à 0.70) devraient être retenues pour effectuer l'apprentissage final et la classification. Cette dernière approche a été abandonnée car nous avons rencontré trois problèmes majeurs. Premièrement, nous obtenions souvent des poids trop bas (entre -0.3 et +0.3). Deuxièmement, pour chaque catégorie d'information le RNA fournissait des poids qui varient selon l'algorithme d'entraînement. Et finalement, souvent pour la même catégorie d'information les variables prédictives "sélectionnées" par un algorithme donné sont en contradiction avec les variables prédictives sélectionnées par un autre algorithme numérique. Ce dernier constat amène à une difficulté majeure au niveau de la compréhension et l'interprétation des modèles en cas d'une implémentation réelle (industrielle) dans le cadre de la prédiction boursière et gestion de portefeuille.

d'autre part quel type d'algorithme numérique d'apprentissage permet au réseau de neurones de fournir la prévision la plus précise possible.

Le marché américain est choisi parce qu'il est un objet d'étude et de validation populaire auprès de la communauté académique (Atsalakis et Valavanis, 2009; Yu et al., 2009; Chang et al., 2011; Huang, 2011; Wang et al., 2011), ce qui permet de comparer nos résultats à ceux de la littérature.

Finalement, nous rappelons que nous ne faisons pas une comparaison entre modèles statistiques et modèles de réseaux de neurones artificiels. Nous utilisons les techniques statistiques pour prétraiter les données et d'y extraire les informations pertinentes à la modélisation par réseaux de neurones et machines à supports de vecteurs.

En particulier, les méthodes statistiques seront employées pour rendre les données stationnaires, modéliser le processus des rendements et extraire sa volatilité, et finalement pour déterminer les variables prédictives. Avant tout, notre thèse de doctorat est principalement une étude empirique et comparative qui requiert par conséquent beaucoup de prétraitements statistiques des données.

CHAPITRE II

MÉTHODOLOGIE

Pour vérifier quel type d'information permet d'obtenir les meilleures prévisions, quel algorithme numérique est plus performant, et si la combinaison de toutes les informations améliore la prédiction; nous avons conçu les systèmes prédictifs présentés dans la section 2.1. L'analyse par ondelette discrète (ordinaire) et l'analyse par paquets d'ondelettes (nouvelle génération) sont introduites au niveau de la section 2.2. Le réseau de neurones artificiels est présenté dans la section 2.3. Les algorithmes (méthodes) numériques sont présentés dans la section 2.4. Les machines à support de vecteurs sont introduites au niveau de la section 2.5. Finalement, les mesures de performances sont présentées au niveau de la section 2.6.

2.1 Conception des expérimentations

Le réseau de neurones artificiels (RNA) est l'outil de l'intelligence artificielle computationnelle le plus utilisé dans le domaine de la prédiction boursière (Atsalakis et Kimon, 2009) à cause (a) qu'il est non-paramétrique, non-linéaire et bien adapté à la modélisation des séries temporelles, (b) il ne requière aucune hypothèse relative au processus qui génère les données ou la nature de leur distribution, (c) il est moins susceptible au problème de mauvaise spécification du modèle que les modèles statistiques paramétriques, (d) il est capable de réaliser une projection entré-sortie et ce même lorsque la vraie nature de la relation entre l'entrée et la sortie est inconnue, (e) il est capable d'approximer n'importe quelle fonction mesurable et continue avec une erreur arbitraire prédéfini (Cybenko, 1989; Hornik et al., 1989), et (f) il est

adaptatif ce qui implique que le pouvoir de généralisation demeure performant, stable et robuste dans un environnement non stationnaire (Khashei et Bijari, 2010).

Le choix des variables prédictives est d'une importance extrême pour la modélisation et la prédiction des séries temporelles boursières car il détermine l'interprétation des résultats et le niveau de la performance des prévisions.

Il existe en général quatre types d'information pour prédire le marché boursier: économique, technique, psychologique, et l'information historique obtenue par analyse statistique de la série temporelle de l'indice boursier à prévoir.

L'information économique est composée des données et des rapports sur l'économie qui sont publiés à des intervalles réguliers (quotidien, mensuel ou trimestriel) selon la nature de l'information. L'information technique est composée d'indicateurs de l'analyse technique pour décrire l'activité du marché. L'information statistique est fondée sur l'historique des rendements pour détecter une éventuelle longue ou courte mémoire dans la série temporelle. En d'autres mots, il s'agit de la modélisation de la dynamique interne de la série en question. L'information psychologique contient des biais psychologiques pour expliquer le comportement de l'investisseur (marché) et donc prédire le marché.

Pour sélectionner les variables les plus pertinentes dans chaque catégorie d'information (économique, technique, sentiment) au RNA, nous allons employer le test de causalité de Granger (Granger, 1969) qui permet d'identifier les entrées qui montrent un lien causal au sens de Granger avec le marché. En d'autres mots, seules les entrées qui influencent statistiquement et significativement le comportement de l'indice boursier seront retenues pour l'apprentissage du RNA. Le test de causalité de Granger (1969) est présenté au prochain chapitre qui porte sur les données et prétraitement.

Pour déterminer quel type d'information permet de mieux prédire les tendances futures du marché boursier nous allons utiliser quatre RNA chacun avec une seule catégorie d'information. Ensuite, nous emploierons un RNA qui utilise toutes les catégories d'information pour examiner l'impact de leur combinaison sur la qualité des prévisions obtenues. Finalement, nous allons décomposer la série temporelles en employant de l'analyse multi-résolution afin d'obtenir de l'information fréquentielle qui sera employée à son tour pour prédire la tendance future du marché boursier. Ce type d'information a l'avantage de caractériser la série temporelle en termes de tendances à long terme (approximation) et de variations locales (détails), des informations qui sont difficilement obtenues par d'autres techniques telles la transformée de Fourier.

En particulier, nous allons utiliser deux types d'informations extraites par analyse multi-résolution à présenter au RNA; à savoir de l'information extraite par analyse en ondelettes discrète (AOD) et de l'information extraite par analyse en paquets d'ondelettes (APO). L'AOD est l'approche standard ou ordinaire (Charles, 1992; Mallat, 1999) qui effectue une décomposition du signal (série temporelle) seulement à partir de ses composantes de basses fréquences; par contre l'APO est la nouvelle génération d'analyse par ondelettes qui permet d'effectuer une meilleure analyse du signal en considérant toutes ses sous-bandes (Coifman et Wickerhauser, 1992).

Pour des fins de comparaisons au RNA, les mêmes expériences seront effectuées en employant le SVM comme classificateurs. En effet, le SVM constitue le classificateur de référence.

Pour prédire la tendance¹⁷ à la hausse ou à la baisse de l'indice boursier S&P500, alors on devrait approximer la fonction suivante :

$$y_t = f(x_t) \quad (2.1)$$

Où t est l'indice de temps, x est la variable d'entrée ou variable prédictive, et y prend la valeur +1 lorsque l'indice boursier varie à la hausse et prend la valeur -1 lorsque l'indice boursier varie à la baisse. La variable de sortie y est définie comme suit :

$$y_t = +1 \quad \text{si} \quad R_t > 0 \quad (2.2)$$

$$y_t = -1 \quad \text{si} \quad R_t < 0 \quad (2.3)$$

Où, R est le taux de variation de l'indice boursier I . Ce dernier est défini par la suite :

$$R_t = \frac{\Delta I_t}{I_t} = \frac{I_t - I_{t-1}}{I_{t-1}} \quad (2.4)$$

Lorsque les observations sont fréquentes, et donc les rendements petits en général, on lui préfère la forme suivante, dite continue ou log-normale¹⁸ (Carmona, 2003) :

$$R_t = \log\left(\frac{I_t}{I_{t-1}}\right) \quad (2.5)$$

La forme de l'équation 2.5 est largement utilisée en finance empirique dans le cas d'observations fréquentes de la variable d'intérêt. Lorsque les données sont observées

¹⁷ Nous avons vérifié empiriquement à l'avance que l'indice boursier varie toujours soit à la hausse ou soit à la baisse, et donc sa tendance (variation) n'est jamais égale à zéro. De ce fait, notre problème de classification revient à un problème de détection de deux classes seulement et non pas trois classes; la troisième étant par exemple $y = 0$ lorsque $R = 0$.

¹⁸ La forme provient de l'approximation $\log(x) = (x-1)$ au voisinage de 1.

de façon hebdomadaire, mensuelle ou trimestrielle, la forme discrète est privilégiée. Dans cette thèse, la forme continue est utilisée.

Les expériences à réaliser sont montrées dans la Figure 2.1 où R , VM , IT , MS , représentent respectivement la variation de l'indice boursier, les variables macroéconomiques, les indicateurs techniques et les mesures de sentiment. Par ailleurs, $AODSA$, $APOSA$, et $APOSHF$ sont respectivement le signal d'approximation obtenu par l'analyse en ondelette discrète (AOD), le signal d'approximation obtenu par l'analyse en paquet d'ondelettes (APO), et le signal de haute fréquence obtenu par APO. Finalement, La variable t indique le temps. Le terme "causal" veut dire que seules les variables prédictives qui montrent un lien causal - au sens de Granger (1969) - avec la série temporelle $R(t)$ sont retenues.

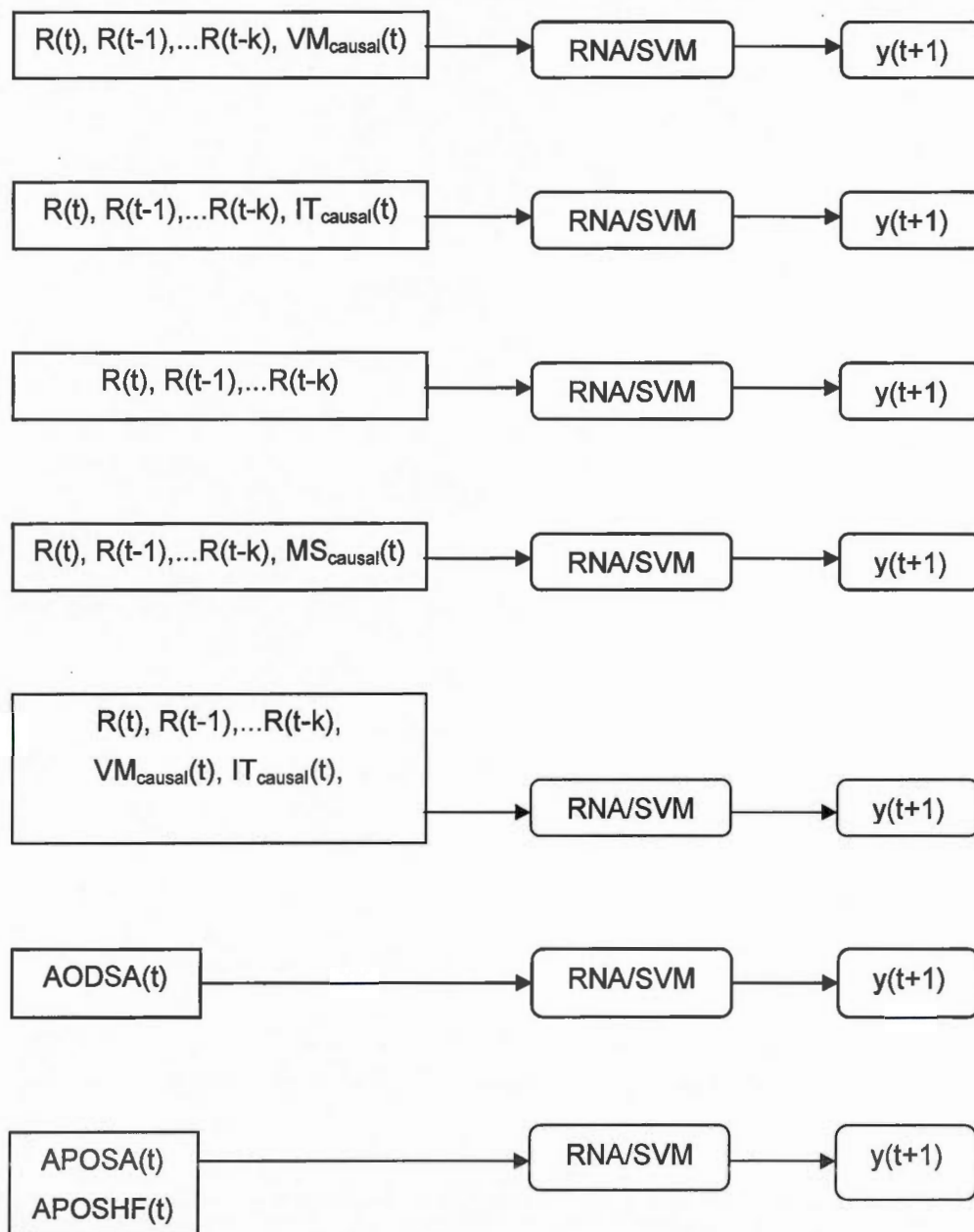


Figure 2.1 Systèmes de prédiction

2.2 La transformée en ondelette

Au niveau de cette section, nous allons présenter deux types d'informations extraites par analyse multi-résolution; à savoir de l'information extraite par analyse en ondelettes discrète (AOD) (Charles, 1992; Mallat, 1999) et de l'information extraite par analyse en paquets d'ondelettes (APO) (Coifman et Wickerhauser, 1992).

2.2.1 La transformée ordinaire en ondelette discrète

L'analyse multi-résolution par ondelettes est un outil mathématique qui sert à transformer des données dans le domaine du temps à plusieurs et différentes "couches" dans le domaine de fréquence. L'analyse par ondelette est supérieure à l'analyse de Fourier car elle localise de l'information simultanément dans le domaine du temps et dans le domaine des fréquences. Ceci permet une meilleure analyse des données observées dans différentes échelles. En particulier, l'analyse de Fourier utilise la somme des fonctions de sinus et de cosinus dans différentes longueurs d'ondes pour représenter une fonction donnée. Cependant, les fonctions sinus et cosinus sont des fonctions périodiques et ne sont pas locales (allant de moins l'infini à plus l'infini). Par contre, l'analyse par ondelette (AO) est localisée à la fois dans le temps et dans l'échelle. De plus, l'AO décompose en plusieurs composantes fréquentielles pour effectuer des analyses individuelles selon une échelle prédéterminée. Par exemple, à haute échelle l'AO permet de concentrer l'analyse sur les détails comme les sommets et les points de transitions (passage d'une période à une autre); à petite échelle, l'AO capture les tendances à long terme. Par conséquent, l'AO est un bon candidat pour l'analyse des séries temporelles –notamment les séries des rendements boursiers- en capturant séparément et simultanément les tendances à long terme (dynamique structurelle) et les détails dans le domaine fréquentiel qui contiennent de l'information plus précise qui se trouve être cachée dans la série temporelle originale. En somme, l'utilisation simultanée de ces deux types d'information peut permettre d'améliorer les prévisions des séries temporelles par

des systèmes intelligents comme les RNA. De plus, plusieurs niveaux de détails sont possibles. Par exemple, on peut analyser un signal S par ondelette au niveau 3 (défini plus loin) pour obtenir la décomposition suivante :

$$S = A_1 + D_1 = (A_2 + D_2) + D_1 = (A_3 + D_3) + D_2 + D_1 \quad (2.6)$$

Où A est le signal approximé et D est le signal détaillé, les indices 1,2, et 3 référant au niveau d'analyse. La Figure 2.2 montre l'arbre de décomposition de cet exemple.

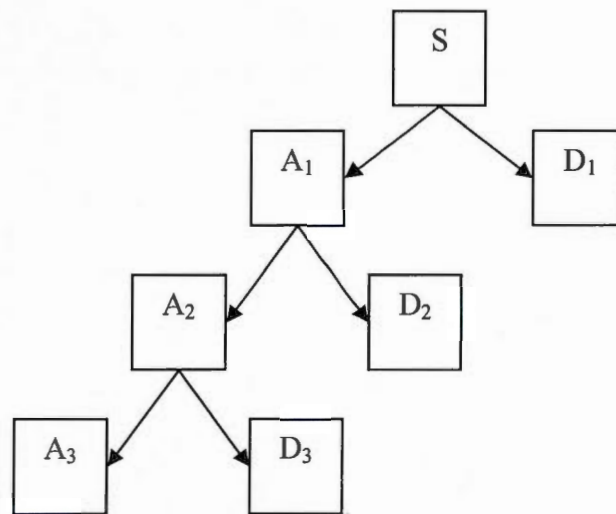


Figure 2.2 Arbre de décomposition par ondelette ordinaire

En termes mathématiques, la transformée en ondelettes continue d'un signal $s(t)$ est définie comme suit (Strang et Nguyen, 1996; Zhang et al. 2010; zhang et al., 2011):

$$W_{\psi}(a,b) = \int_{-\infty}^{+\infty} s(t) \psi_{a,b}^*(t) dt \quad (2.7)$$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \quad (2.8)$$

Où $\psi_{a,b}()$ est une fonction d'ondelette mère donnée, et a et b sont les paramètres de dilation et de translation respectivement.

La transformée en ondelettes discrète est obtenue par échantillonnage de a et b . dans le cas d'un échantillonnage dyadique les paramètres a et b sont respectivement des puissances du chiffre 2. Ainsi, les coefficients d'ondelette correspondants sont donnés par :

$$c_{jk} = W_\psi(2^{-j}, 2^{-j}k) \quad (2.9)$$

Où j représente le facteur de dilation discrète et k représente le facteur de translation discrète. En termes pratiques, l'équation (1) est essentiellement une mesure de corrélation entre le signal et différentes ondelettes dérivées d'une ondelette mère.

En pratique, l'équation précédente est impossible à mettre en œuvre sans troncation des indices, ce qui remettrait en cause son utilité. De ce fait, la décomposition par ondelette discrète est obtenue plutôt par une séquence de filtrages pass-haut et pass-bas (Mallat, 1989), reflétant ainsi sa capacité d'effectuer une analyse d'approximation et de détail simultanément. La forme discrète de l'équation (1) est alors comme suit (Zhang et al. 2010; zhang et al., 2011) :

$$\begin{aligned} ca_{j,k}[x(t)] &= DS \left[\sum x(t) g_j^*(t - 2^j k) \right] \\ cd_{j,k}[x(t)] &= DS \left[\sum x(t) h_j^*(t - 2^j k) \right] \end{aligned} \quad (2.10)$$

Où les coefficients $ca_{j,k}$ et $cd_{j,k}$ représentent les composantes d'approximation et de détails, obtenues respectivement par le filtre pass-bas $g(n)$ et le filtre pass-haut $h(n)$. Le terme DS représente l'opérateur de décimation par un facteur de 2.

Pour illustrer l'apport de l'analyse par ondelette discrète, nous avons simulé un signal en pas d'escalier corrompu avec du bruit gaussien. Ensuite, nous avons analysé ce signal en utilisant une ondelette de type Symlet au premier niveau de décomposition. La Figure 2.3 montre le signal original en pas d'escalier corrompu avec du bruit gaussien, sa décomposition d'approximation, et sa décomposition de détails.

Cet exemple montre la capacité de l'analyse par ondelette à décomposer un signal pour y extraire de l'information pertinente; par exemple les variations locales (détails) et les tendances à long terme (approximation).

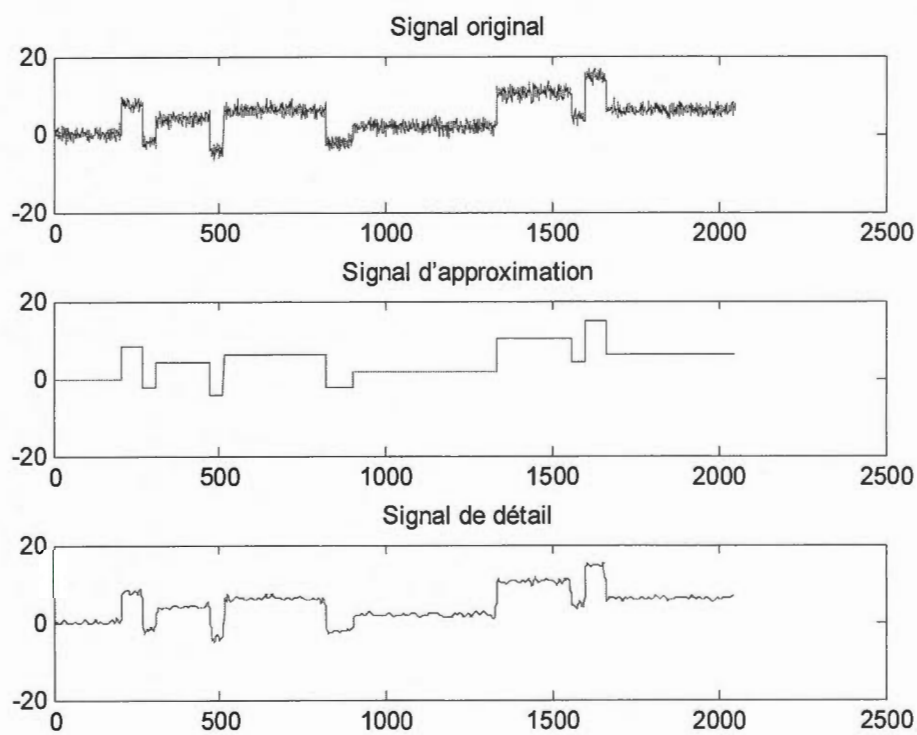


Figure 2.3 Analyse par ondelette discrète d'un signal gaussien à pas d'échelle

La transformée en paquet d'ondelettes est une décomposition de plus des signaux d'approximation (basse-fréquence) et de détails (haute-fréquence) à chaque niveau de décomposition. De ce fait, une meilleure résolution du signal initial est obtenue dans le domaine temps-fréquence. Dans notre thèse, nous considérons l'analyse en paquet d'ondelettes (APO) qui est une généralisation de la décomposition par ondelette dans la mesure où elle offre de plus larges possibilités pour l'analyse du signal (S). Dans notre cas, le signal réfère aux séries temporelles.

Dans l'analyse par ondelette, le signal est décomposé en sous signal approximé (approximation : A) et sous signal détaillé (détail : D). Le signal approximé est à son tour décomposé à un deuxième niveau en approximation et en détails. Ce processus

est encore répété. Par exemple, pour n-niveau de décomposition, il y a n+1 possibilités pour la décomposition du signal.

Contrairement à l'AO, dans l'APO les décompositions en approximation et en détails sont à leurs tours décomposées; ayant pour résultat 2^{2^n-1} différents chemins pour décomposer le signal S. Par exemple, on peut analyser un signal S par APO au niveau 3 et on obtient l'arbre présenté dans la Figure 2.4.

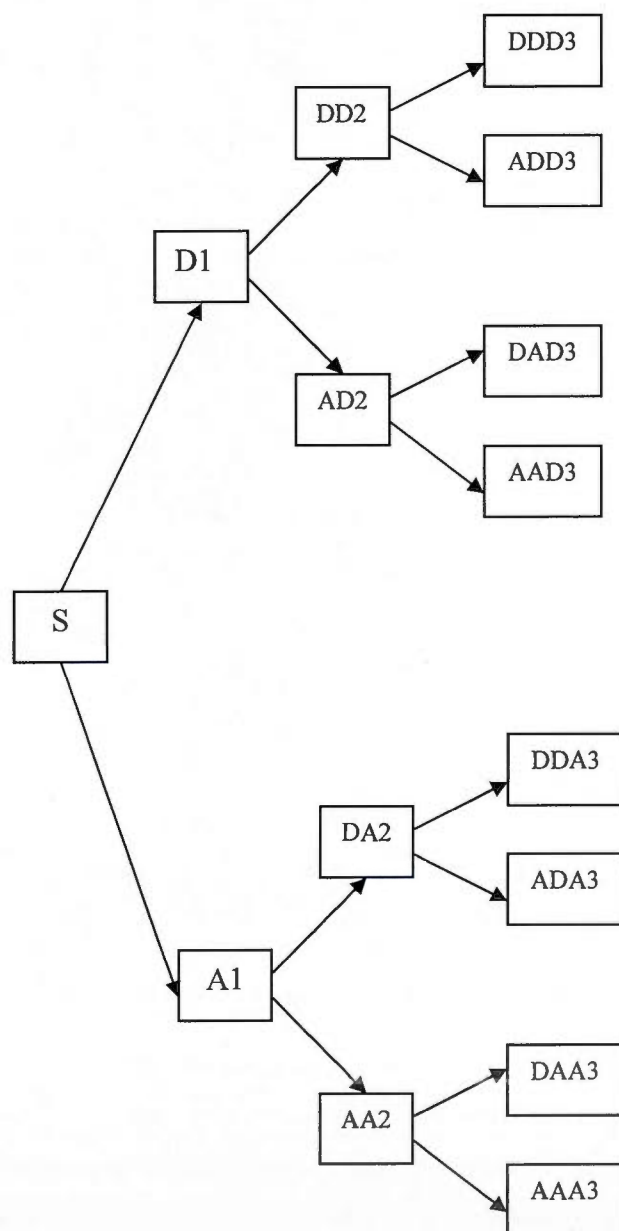


Figure 2.4 Arbre de décomposition par transformée en paquet d'ondelette

En termes mathématiques et selon la notation de Zhao et Ye (2010), la décomposition du signal $s(t)$ par analyse en paquet d'ondelettes (APO) est donnée comme suit:

$$s_p^{n,j} = 2^{-j/2} \int_R s(t) \mu_n(2^{-j}t - p) dt \quad (2.11)$$

Où,

$$0 \leq j \leq S \quad \text{et} \quad 0 \leq n \leq 2^S - 1 \quad (2.12)$$

Où, $\mu_n(t)$ est la fonction ondelette en paquet, j est le niveau de décomposition, p est le paramètre de position de la fonction, n est le numéro du chemin, et S est le niveau de décomposition maximal. Après la décomposition du signal $s(t)$ par la transformée en paquet d'ondelettes, on obtient 2^S séquences au niveau S de décomposition. L'algorithme de décomposition rapide de la transformée en paquet d'ondelettes est donnée comme suit :

$$\left\{ \begin{array}{l} s_k^{2n,j+1} = \sum_{p \in \mathbb{Z}} h(p-2k) s_p^{n,j} \\ s_k^{2n+1,j+1} = \sum_{p \in \mathbb{Z}} g(p-2k) s_p^{n,j} \end{array} \right\} \quad (2.13)$$

Où, $h(i)$ et $g(i)$ sont respectivement les coefficients des filtres de l'ondelette en miroir de quadrature; notamment les filtres passe-haut et passe-bas respectivement.

Pour déterminer la meilleure décomposition (représentation); par exemple trouver le meilleur sous-arbre; la littérature sur les APO (Coifman et al., 1990; Coifman et Wickerhauser, 1992; Ramchandran and M. Vetterli, 1993) propose différents critères de sélection (ou fonctions de coût) qui permettent de déterminer le meilleur chemin qui forme le sous-arbre optimal en se basant sur la quantification du gain d'information à chaque nœud. L'objectif est de choisir le sous-arbre qui minimise la fonction de coût. Il existe trois critères de sélection (fonction de coût) qui sont les

plus utilisés dans la littérature sur l'analyse des signaux par paquet d'ondelettes (Yang et Xu, 2006); incluant la mesure d'entropie de Shannon (Shannon, 1948), la L1-norm, et le taux de distorsion (Ramchandran et Vetterli, 1993).

Ils sont présentés dans le Tableau 2.1 où CS est le critère de sélection, c est le coefficient obtenu par analyse en paquet d'ondelettes (APO), D est le taux de distorsion (quantification¹⁹) du signal (coefficients de l'APO), R est le débit binaire qui mesure de la quantité de données numériques transmises par unité de temps, λ est le facteur de qualité, et i est l'indice de temps.

Tableau 2.1

Critères de sélection de la meilleure décomposition par APO

Critère de sélection ou fonction de coût	Formulation mathématique
Shannon	$CS = -\sum_i c_i^2(t) \log(c_i^2(t))$
L1-Norm	$CS_{norm} = \sum_i c_i $
Taux de distorsion	$CS_{RD}(\lambda) = D + \lambda.R$

Le désavantage du CS de type L1-Norm est le fait qu'il engendre beaucoup de décompositions non nécessaires (Yang et Xu, 2006). Par ailleurs, le principal désavantage du CS de type taux de distorsion est le taux de complexité du calcul relativement haut dû au processus d'optimisation du paramètre λ (Ramchandran et Vetterli, 1993; Yang et Xu, 2006). De ce fait, nous considérons le critère d'information de Shannon (1948). Ce dernier est basé sur la théorie de l'information

¹⁹ La quantification est une technique largement utilisée en traitement de signal qui sert à approximer un signal continu par des valeurs discrète.

ce qui permet de caractériser le niveau aléatoire d'une série temporelle et de quantifier la différence entre deux distributions de probabilités; par exemple entre deux chemins de décomposition. De plus, le critère d'entropie de Shannon est plus adapté pour le choix optimal des signaux décomposés par l'APO (Coifman et Wickerhauser, 1992; Ruiz et al., 1999; Wu et Liu, 2009).

Pour montrer le potentiel de l'APO dans le traitement des signaux aléatoire, un exemple de dé-bruitage par l'APO d'un signal corrompu avec du bruit gaussien est montré dans la Figure 2.5. En effet, l'APO a été appliquée avec succès dans les problèmes de dé-bruitage de signaux unidimensionnels et bidimensionnels (Shui et al., 2007; Ye et al., 2009; Kubinyi et al., 2011; Zhou et al., 2011). Cet exemple montre clairement la capacité de l'APO à isoler le signal original du bruit gaussien.

Un exemple de décomposition du signal (le rendement du marché boursier) en composantes d'approximations et de détails par l'APO est présenté au niveau du Chapitre 3 (Données et Prétraitements).

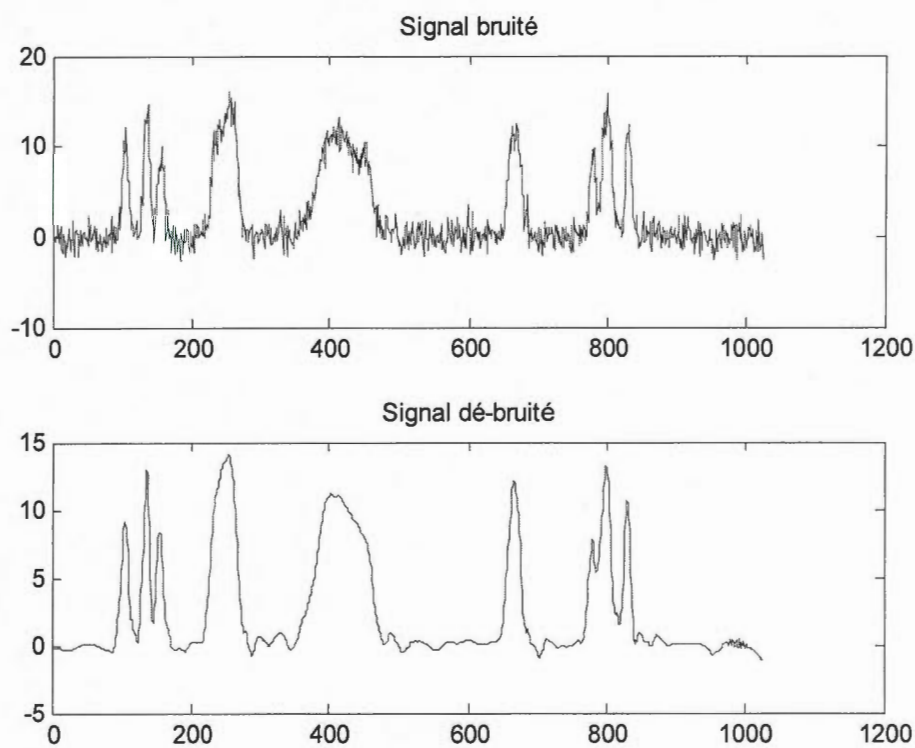


Figure 2.5 Dé-bruitage d'un signal par la transformée en paquet d'ondelette

2.2.2 Le choix du type d'ondelette

Il existe plusieurs types d'ondelettes dans la littérature (exemple : Haar, Daubechies, Symlet, etc.) et chacune d'elles possède ses propres caractéristiques (Daubechies, 1994). Parmi les caractéristiques les plus importantes, on peut citer la symétrie, le nombre de moments²⁰ nuls (ordre), la régularité, l'orthogonalité, et si l'ondelette en question existe en forme discrète, continue, ou les deux. La symétrie est surtout utile pour réduire les artefacts lorsqu'on compresse une image en effectuant un lissage sur les bords de l'image.

²⁰ On dit qu'une ondelette ψ possède un moment d'ordre k si $\int_{-\infty}^{+\infty} t^k \psi(t) dt = 0$.

Le nombre de moments nuls renseigne sur la précision avec laquelle l'ondelette pourra capturer les détails du signal. Plus le nombre de moments nuls sera grand plus la précision sera meilleure; et donc une meilleure analyse du signal. L'orthogonalité réfère à la relation (orthogonalité) entre le signal d'approximation et le signal de détail. Par conséquent, elle est associée à l'erreur et donc renseigne sur la qualité d'approximation.

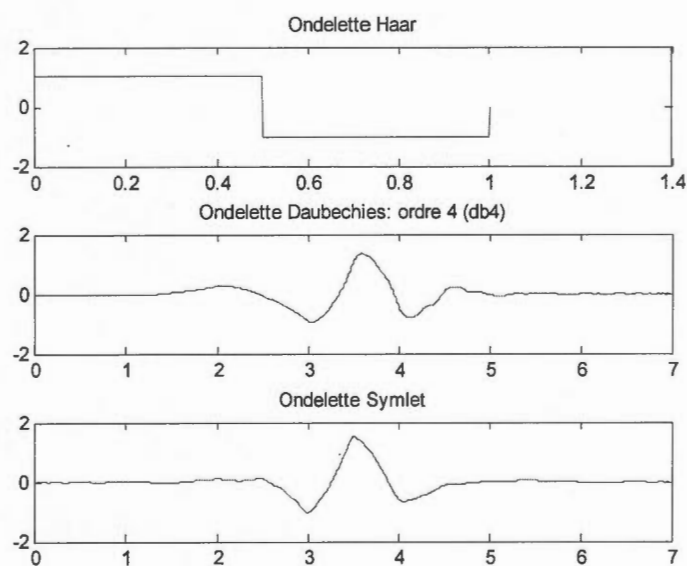
La régularité permet d'améliorer la qualité de la reconstruction. De ce fait, elle est importante pour les applications de compression. En d'autres mots, elle rend l'approximation meilleure. La régularité est importante dans la mesure où elle permet d'obtenir des caractéristiques pertinentes du signal ou d'une image. Elle est aussi utile pour l'estimation d'une fonction non-linéaire par régression. Finalement, l'ondelette peut avoir une représentation discrète ou continue.

La forme la plus utilisée est la forme discrète car elle est facile à implémenter. Cependant, la forme continue devient de plus en plus utilisée dans certaines applications biomédicales comme l'analyse des signaux cérébraux unidimensionnels pour la détection précise des anomalies. Plusieurs ondelettes (exemple : Meyer, Gaussienne, Mexicaine, Coiflet, etc.) et caractéristiques (exemples : support compact, analyse biorthogonale, forme explicite, etc.) sont présentées et discutées rigoureusement dans Daubechies (1994). Le Tableau 2.2 présente une comparaison sommaire des ondelettes Haar et Daubechies (db4) utilisées pour l'analyse et la prédiction boursière (Huang et Wu, 2010, Hsieh et al., 2011; Huang, 2011) par rapport à l'ondelette Symlet d'ordre deux. Par ailleurs, les trois ondelettes sont montrées dans la Figure 2.6.

Tableau 2.2

Comparaisons des ondelettes Haar, Daubechies, et Symlet

	Haar	Daubechies	Symlet
Symétrie	oui	asymétrique	Quasi-symétrique
Nombre de moments	Un seul (basique)	Arbitraire (2,3,...)	Arbitraire (2,3,...)
Régularité	discontinue	Arbitraire (continue)	Arbitraire (continue)
Orthogonalité	oui	oui	oui
Forme discrète	oui	oui	oui
Forme continue	oui	oui	oui

**Figure 2.6** Les ondelettes Haar, Daubechies, et Symlet

La fonction d'ondelette adoptée pour effectuer l'analyse multi-résolution est la Symlet d'ordre 2 (Daubechies, 1992), car outre qu'elle possède les mêmes caractéristiques (notamment la régularité) que l'ondelette bien connue de Daubechies (Daubechies, 1992); elle permet d'avoir un filtre presque symétrique. Le niveau de décomposition est arbitrairement établi à 3. Comme les séries temporelles étudiées sont stochastiques, nous faisons l'hypothèse qu'une fonction ondelette presque symétrique (et non pas totalement symétrique) capturerait mieux les irrégularités derrière le processus qui gouverne les séries. Quand à la régularité, elle permet d'obtenir une meilleure approximation. Dans notre cas, la régularité est très utile puisque nous employons des systèmes intelligents capables d'approximer des fonctions non-linéaires.

2.3 Le réseau de neurones artificiels

Le réseau de neurones artificiels (RNA) a l'avantage d'être simple et d'être capable d'approximer n'importe quelle fonction non linéaire (Lippman, 1987 ; Baum, 1988; Cybenko, 1989; Hornik, 1991). Le réseau de neurones artificiels (RNA) de type perceptron multicouches (PMC) (Rumelhart et al, 1986) est composé d'une couche d'entrée, une ou plusieurs couches cachées, et d'une couche de sortie.

Les neurones de la couche d'entrée reçoivent et distribuent les signaux (x_i) d'entrée aux neurones de la couche cachée. Chaque neurone j dans la couche cachée reçoit tout les autres signaux d'entrée (x_i) pondérés avec des poids de connections w_{ji} . Finalement, chaque neurone j dans la couche cachée calcule sa sortie y_j comme étant une fonction f_a de la somme des signaux d'entrées qui sont pondérés comme suit :

$$y_j = f_a\left(\sum w_{ji}x_i\right) \quad (2.14)$$

Où f_a est une fonction d'activation. La méthode de la rétro-propagation d'erreur (Rumelhart et al, 1986) est largement utilisée pour faire l'apprentissage de réseaux de neurones non récurrents. L'erreur calculée en sortie du réseau est rétro-propagée vers

les couches cachées. L'objectif est d'adapter les poids du réseau par la descente du gradient (voir section suivante) et la propagation de l'erreur en minimisant la fonction de coût suivante :

$$E = \frac{1}{2} \sum (d - y)^2 = \frac{1}{2} \sum e^2 \quad (2.15)$$

Où, y et d sont les sorties actuelle et désirée respectivement dans chaque nœud. L'entraînement du réseau de neurones consiste à ajuster les poids du réseau en employant différents algorithmes. L'algorithme d'entraînement détermine $\Delta w_{ji}(t)$ qui est le changement du poids entre le neurone i et le neurone j au temps t . Les poids sont mis à jour selon l'équation suivante :

$$w_{ji}(t+1) = w_{ji}(t) + \Delta w_{ji}(t) \quad (2.16)$$

Ou encore en termes de notation allégée :

$$w_{i+1} = w_i + \Delta w_i \quad (2.17)$$

La fonction d'activation considérée est la tangente hyperbolique dont les sorties appartiennent à l'intervalle $[-1,1]$. Ce choix est bien adapté à notre problème de classification : hausse (+1) versus baisse (-1). Elle est définie comme suit :

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}, \quad (2.18)$$

Par ailleurs, la fonction de sortie est une "pureline" (). La Figure 2.7 présente une architecture d'un RNA avec deux neurones dans la couche d'entrée, quatre neurones dans la couche cachée et un neurone dans la couche de sortie. Les poids et les biais des neurones sont initialisés de manière aléatoire entre -0.1 et 0.1 au voisinage de zéro.

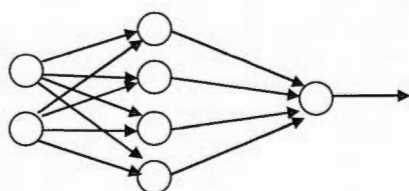


Figure 2.7 Un exemple d'architecture de RNA

Au niveau de la topologie du RNA à employer, le nombre de neurones de la couche d'entrée sera fixé au nombre des variables prédictives, et le nombre de neurones dans la couche cachée sera fixé au même nombre, au double et au triple de neurones dans la couche d'entrée. En fait, il n'existe pas de règles formelles pour déterminer le nombre de neurones et de couches cachées dans un RNA (Burkitt, 1991; Kavzoglu, 1999; Shafi et al., 2006). Ces différents nombres de neurones dans la couche cachée sont choisis, car un grand nombre de neurones dans la couche cachée rend le(s) neurone(s) de la couche de sortie instable(s) (Shibata, 2009). Ce résultat a été déjà constaté par Shafi et al., (2006). Similairement, lorsque le nombre de neurones dans la couche cachée est petit alors le(s) neurone(s) de la couche de sortie devient aussi instable(s) (Shibata, 2009).

Ainsi, nous avons choisi de fixer le nombre de neurones dans la couche cachée au même nombre, au double et au triple de neurones dans la couche d'entrée.

Le nombre des couches cachées sera fixé à un puisque le RNA standard est un approximateur universel de fonctions (Cybenko, 1989; Funahashi, 1989; Hornik, 1991). De plus, la décision concernant le nombre de couche cachée dans le réseau relève du domaine de l'art et de l'expérience de l'utilisateur (Burkitt, 1991); et donc il n'existe pas une règle pour déterminer le nombre de neurones et des couches cachées (Kavzoglu, 1999; Shafi et al., 2006).

En outre, la complexité du calcul grandit avec la taille croissante du réseau; et il est donc plus efficace de garder la taille du réseau la plus petite possible (Burkitt, 1991). D'ailleurs, dans le cadre d'une modélisation d'un système complexe et non-linéaire (séries temporelles bruitées de spectrogramme) Shafi et al., (2006) ont conclu qu'il n'y a pas un besoin pour plus d'une couche cachée dans le réseau de neurones car la performance de ce dernier se détériore. En d'autres mots, l'erreur du RNA augmente avec le nombre de couches cachées.

Dans ce sens, pour la plupart des problèmes pratiques il n'y a pas de raisons pour mettre plus d'une seule couche cachée dans le RNA (Panchal et al., 2011). Finalement, les résultats théoriques de Fausett (1994) ont montré qu'une seule couche cachée dans le RNA à rétro-propagation d'erreur est suffisante pour approximer n'importe quelle fonction continue à un degré arbitraire de précision.

Cependant, pour le meilleur modèle obtenu (le couple algorithme numérique et information qui obtient la meilleure performance de classification) nous allons faire varier le nombre de couches cachées pour vérifier s'il y a possibilité d'améliorer davantage sa performance obtenue avec une seule couche cachée.

2.4 Les méthodes d'optimisation numérique

L'algorithme le plus connu pour l'optimisation des poids du RNA et l'algorithme standard de la retro-propagation qui utilise la descente du gradient (voir Section 2.4.1) où les poids du réseau varient dans la direction négative du gradient de la fonction de performance (fonction de coût). Cependant, l'algorithme de la descente du gradient est souvent lent pour converger. Il existe différentes approches pour accélérer la convergence de la descente du gradient et donc du réseau (système). Ils peuvent être divisés en trois catégories (Haykin, 1994; Principe et al., 1999; Bose et Liang, 2001).

La première utilise des techniques heuristiques, la deuxième catégorie utilise les méthodes standards d'optimisation numérique, et la troisième catégorie utilise les approches d'optimisation stochastique.

Les approches de la première catégorie sont développées sur la base de la performance de l'algorithme standard de la descente du gradient. L'objectif est d'améliorer la vitesse de convergence du réseau (système) en adaptant le taux d'apprentissage de la descente du gradient.

Par contre, les approches de la deuxième catégorie utilisent les techniques d'optimisation numérique pour accélérer la convergence du réseau. Finalement, les approches stochastiques sont basées sur la minimisation de la fonction objective en accumulant de l'information sur la distribution des exemples présentés au RNA lors de l'entraînement. En effet, elles approximent une relation probabiliste entre les entrées et les sorties. Les techniques heuristiques comprennent la descente du gradient, la descente du gradient avec momentum, la descente du gradient avec taux d'apprentissage variable, et l'algorithme de la retro-propagation résiliente. Les algorithmes numériques comprennent, le gradient conjugué, le quasi-Newton, et le Levenberg-Marquard (L-M).

Finalement, parmi les méthodes d'optimisation stochastique on peut citer le modèle "Search and Converge" de Darken et al., (1992) et le "Effective Learning Rate" de Leen et Orr(1994).

Le principal désavantage des approches heuristiques réside dans le nombre important des paramètres à prédéterminer et aussi dans la difficulté à trouver leurs valeurs optimales et aussi dans leur caractère aléatoire. De plus, la vitesse de la convergence n'est pas toujours améliorée (Moreira et Fiesler, 1995; Amini, 2008). Finalement, elles ont l'inconvénient d'intégrer une recherche aléatoire (Gili et Winker, 2008).

De leur côté, les méthodes d'optimisation stochastique introduisent du bruit dans le réseau pendant l'entraînement à cause de leur nature aléatoire et consomment souvent plus de temps pour converger (Moreira et Fiesler, 1995). Par contre, les approches numériques comme le gradient conjugué permettent effectivement d'obtenir une bonne approximation de la fonction à étudier et aussi une convergence du système plus rapide comparativement à l'approche standard de la descente du gradient (Sharma et al., 2007; Noori et al., 2010; Hsu et al., 2011). De plus, elles requièrent le moins de paramètres à prédéterminer parmi les trois catégories et requièrent seulement l'hypothèse que la fonction à approximer est continue et est deux fois dérivable.

De ce fait, les approches numériques sont les plus utilisées en finance (Atsalakis et Valavanis, 2009; Wang et al., 2011); en particulier la technique de Levenberg-Marquard (L-M) (Nocedal et Wright, 2006; Eligius et Boglárka, 2010). Par conséquence, nous allons étudier l'impact des méthodes d'optimisation numériques sur la performance du RNA dans le cadre de la prédiction de la tendance future du marché boursier. En suivant les notations de (Sharma et al., 2007; Noori et al., 2010; Charniya et al., 2011), les sections 2.4.1 à 2.4.4 présentent la méthodes de la descente du gradient et les méthodes numériques employées.

2.4.1 La méthode de la descente du gradient

L'Algorithme standard de la retro-propagation utilise l'algorithme de la descente du gradient dans lequel les poids du réseau varient dans la direction négative du gradient de la fonction de performance (Sharma et al., 2007; Noori et al., 2010; Charniya et al., 2011) :

$$w_{t+1} = w_t - \alpha_t \cdot g_t \quad (2.19)$$

Où w est le vecteur de poids, g et le gradient, α est le taux d'apprentissage, et t est l'indice de temps (itération). Le gradient g est donné comme suit :

$$g_t = \nabla E(w_t) = J(w_t)^T E(w_t) \quad (2.20)$$

Où $J(w)$ est le Jacobien (matrice des dérivées premières) de la fonction $E(w)$. Malgré le fait que la fonction de performance diminue rapidement, cet algorithme ne fournit pas nécessairement la convergence la plus rapide. Ainsi, d'autres algorithmes numériques ont été proposés pour accélérer la convergence comme la méthode du gradient conjugué, la méthode quasi-Newton de BFGS, et la méthode de Levenberg-Marquard (L-M). Elles sont présentées dans la section qui suit.

2.4.2 La méthode du gradient conjugué

Dans le cadre de la méthode du gradient conjugué, la recherche de la direction est faite tout au long des directions conjuguées (Hagan et al., 1996). La technique numérique du gradient conjugué profite implicitement de l'information sur le Hessien (H) qui est la matrice des dérivées secondes partielles. Tous les algorithmes numériques de cette catégorie commencent par une recherche dans la direction de la descente du gradient lors de la première itération :

$$p_0 = -\nabla E(w_{t=0}) = -g_0 \quad (2.21)$$

Pour déterminer la distance optimale à parcourir sur la direction de recherché actuelle, une recherche de ligne (Line Search) est effectuée :

$$w_{t+1} = w_t + \alpha_t \cdot p_t \quad (2.22)$$

Ensuite, la prochaine direction de recherche est déterminée de façon à ce qu'elle soit le conjugué de la direction de recherche précédente. Par exemple, deux directions sont considérées conjuguées lorsque :

$$p_i^T H p_j = 0 \quad \text{pour } i \neq j \quad (2.23)$$

Où correspond à la direction d'optimisation à l'étape i . dans le cas où le Hessien est une matrice identité, alors la recherche est effectuée dans deux directions perpendiculaires.

La combinaison de la nouvelle direction de descente avec la direction de recherche précédente est la procédure générale pour déterminer la nouvelle direction de recherche :

$$p_t = -g_t + \beta_t p_{t-1} \quad (2.24)$$

Il existe plusieurs types d'algorithme du gradient conjugué pour calculer le paramètre β . Les plus utilisés sont Polak-Ribiere, Fletcher-Reeves, et Powell-Beale.

2.4.2.1 Polak-Ribière

La mise à jour de la direction de la recherche à chaque itération est déterminée comme suit :

$$\beta_t = \frac{\Delta g_t^T \cdot g_k}{g_{t-1}^T \cdot g_{t-1}} \quad (2.25)$$

Il s'agit du produit scalaire du changement du gradient divisé par la norme au carré du gradient précédent.

2.4.2.2 Fletcher-Reeves

La procédure de mise à jour de la nouvelle direction de recherche selon Fletcher et Reeves (1964) est défini comme suit :

$$\beta_t = \frac{g_t^T \cdot g_k}{g_{t-1}^T \cdot g_{t-1}} \quad (2.26)$$

C'est le ratio de la norme au carré du gradient à l'instant t à la norme au carré du gradient au temps $t-1$.

2.4.2.3 Powell-Beale

Pour tous les algorithmes numériques basés sur le gradient conjugué, la direction de la recherche est initialisée au négatif du gradient. En se basant sur le travail de Beale (1972), et pour améliorer l'efficacité de l'entraînement, Powell (1977) a proposé la condition suivante pour la remise à jour du point de départ de la nouvelle recherche :

$$|g_{i-1}^T g_i| \geq 0.2 \|g_i\|^2 \quad (2.27)$$

Selon cette approche, la mise à jour est faite lorsqu'il existe une petite orthogonalité entre le gradient en cours et le gradient précédent. L'existence de cette petite orthogonalité est testée par la validation de la condition précédente.

2.4.2.4 La méthode Quasi-Newton de BFGS

La méthode de Newton (Nocedal et Wright, 2006; Eligius et Boglárka, 2010) consiste à approximer localement une fonction f par une fonction quadratique et minimise la fonction quadratique approximée à chaque itération t . Autrement dit, il s'agit d'une approximation de second ordre de la fonction f par la série de Taylor comme suit :

$$f(w) \approx f(w_i) + (\nabla f(w_i))^T (w - w_k) + \frac{1}{2} (w - w_k)^T H(w_k) (w - w_k) \quad (2.28)$$

Finalement, cette fonction quadratique est minimisée à :

$$w_i = w_i - H(w_i)^{-1} \nabla E(w_i) \quad (2.29)$$

Dans le cas de la méthode de Newton, la future direction de recherche est donnée comme suit :

$$w_{i+1} = w_i - \alpha_i H_i^{-1} g_k \quad (2.30)$$

Le principe de la méthode quasi-newtonienne BFGS (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) est une généralisation de la méthode de Newton de l'équation. Elle requiert plus de computations à chaque itération et nécessite une plus large mémoire comparativement aux techniques du gradient conjugué car le calcul de l'inverse du hessien est coûteux. En effet, si r est le nombre de poids dans le réseau, alors la taille du Hessien (H) approximé est de l'ordre de $r \times r$. Cependant, il converge généralement très rapidement (Dennis & Schnabel, 1983; Sharma et al., 2007). Le BFGS est donné comme suit :

$$H_{i+1} = H_i + \left(1 + \frac{\Delta g_i^T H_i \Delta g_i}{\Delta g_i^T \Delta w_i} \right) \frac{\Delta w_i \Delta w_i^T}{\Delta w_i^T \Delta g_i} - \frac{A_i \Delta g_i \Delta w_i^T + (A_i \Delta g_i \Delta w_i^T)^T}{\Delta g_i^T \Delta w_i} \quad (2.31)$$

2.4.3 La méthode de Levenberg-Marquard

L'algorithme de Levenberg-Marquard (L-M) (Levenberg, 1944; Marquardt, 1966; Nocedal et Wright, 2006; Eligius et Boglárka, 2010) a été conçu pour accélérer la convergence des modèles de deuxième²¹ ordre sans calculer la matrice des dérivées secondes partielles (le Hessien, H) qui est plus coûteux en termes computationnel. La future direction de recherche est donnée comme suit:

$$w_{i+1} = w_i - [J(w_i)^T J(w_i) + \lambda I]^{-1} J(w_i)^T E \quad (2.32)$$

Où λ est une constante positive et J est le Jacobien qui est la matrice qui contient les premières dérivées des erreurs du réseau par rapports aux poids et E est le vecteur des erreurs du réseau. La valeur de λ est fixée arbitrairement à 0.001.

²¹ Par définition, les techniques numériques de deuxième ordre utilisent le Hessien pour l'approximation de la fonction étudiée.

2.4.4 La détermination du pas de la future direction de recherche

Les méthodes numériques de type gradient conjugué et de type quasi-Newton nécessitent l'élaboration d'une ligne de recherche (Line Search). Il existe plusieurs algorithmes pour effectuer une ligne de recherche; notamment l'approche de Charalambous (1992), la règle d'or (Golden section) de Kiefer (1953), la recherche de Fibonacci (Avriel et Wilde (1966), et Backtracking (Brassard et Bratley, 1995). L'approche de Charalambous (1992) sera appliquée pour le calcul de α_t car elle est une technique hybride conçue spécialement pour l'entraînement des réseaux de neurones. En effet, cet algorithme performe une mise à jour des poids du réseau de façon parallèle et efficiente par rapport aux autres techniques (Charalambous, 1992; Rowan et al., 2007). Cet algorithme utilise conjointement une interpolation cubique et du sectionnement.

De plus, il permet de localiser le minimum plus rapidement que les autres techniques mentionnées précédemment (Charalambous, 1992). Rappelons que l'algorithme standard de la retro-propagation est donné comme suit :

$$w_{t+1} = w_t - \alpha \cdot g_t \quad (2.33)$$

Où α est le taux d'apprentissage constant. La méthode numérique de Charalambous (1992) requiert la détermination de α à chaque itération t qui minimise la fonction suivante à chaque itération :

$$\psi(\alpha_t) = \phi(w_t + \alpha_t \cdot d_t) \quad (2.34)$$

Où, ψ est une fonction quadratique à approximer et :

$$\phi(w_t) = \frac{1}{2} \sum e^2(w_t) \quad (2.35)$$

$$d_t = -\nabla E(w_t) \quad (2.36)$$

Pour atteindre le minimum local, Charalambous (1992) impose deux conditions. La première concerne la borne de droite et la deuxième concerne la borne de gauche. Il s'agit en fait d'un sectionnement qui permet de fournir la valeur optimale de a qui obéit aux deux conditions mentionnées précédemment. La première condition est donnée comme suit :

$$\psi(\alpha_i) \leq \psi(0) + \mu \cdot \alpha_i \cdot \psi'(0) \quad (2.37)$$

Où, $\mu < 0.5$ et,

$$\psi(\alpha) = \frac{\partial \psi(\alpha)}{\partial \alpha} \quad (2.38)$$

La condition de la borne de droite est équivalente à :

$$\phi(w_{i+1}) \leq \phi(w_i) + \mu \cdot \alpha_i \cdot \nabla \phi^T(w_i) \cdot d_i \quad (2.39)$$

La deuxième condition qui concerne la borne de gauche est donnée par :

$$|\psi'(\alpha_i)| \leq -\sigma \cdot \psi'(0) \quad (2.40)$$

Où,

$$\sigma \in (0,1) \text{ et } \sigma \leq \mu \quad (2.41)$$

Elle est équivalente à :

$$|\nabla \phi^T(w_{i+1}) \cdot d_i| \leq -\sigma \cdot \nabla \phi^T(w_i) \cdot d_i \quad (2.42)$$

Finalement, en définissant une fonction cubique pour (α_i) , Charalambous (1992) a montré que la valeur optimale de a est comme suit :

$$\alpha = \max\{\alpha_p, \alpha_q^*\} \quad (2.43)$$

Où, α_p est la valeur de α lors de l'itération précédente et, α_q^* est sa valeur lors de l'actuelle itération, et α_q^* est donnée comme suit:

$$\alpha_q^* = -\frac{2\Delta\phi}{d_i \nabla \phi(w)} \quad (2.44)$$

Dans le cadre de notre thèse, les valeurs de α et de σ sont 0.25 et 0.5 respectivement car elles permettent d'obtenir une meilleur convergence (Charalambous, 1992).

2.5 Les machines à supports de vecteurs

Vapnik (1995) a introduit les machines à vecteurs de support (support vector machines, SVM²²) fondées sur le principe de minimisation du risque structurel, issu de la théorie statistique de l'apprentissage. Le SVM entre dans la catégorie des méthodes à fonctions noyaux. Le pouvoir du SVM est de trouver un compromis entre la capacité d'apprentissage et la capacité de généralisation d'un modèle par apprentissage. De plus, il évite le minimum local sous l'hypothèse que les données sont identiquement et indépendamment distribuées.

Dans cette section, nous allons introduire le principe du SVM dans le cas des données linéairement séparables. Ensuite, nous allons présenter le SVM dans le cas des données non-linéairement séparables.

2.5.1 Le cas des données linéairement séparables

Soit S un ensemble d'exemples $(x_1, y_1), (x_2, y_2), \dots, (x_L, y_L)$, où $x_i \in \mathbb{R}^N$ sont les entrées au système prédictif et $y_i \in \{-1, +1\}$ sont les classes ou les sorties associées aux entrées.

²² SVM sera employé au masculin singulier pour alléger le texte.

À ce niveau, on suppose que les deux classes -1 et +1 sont linéairement séparables. Dans ce cas, on peut trouver un vecteur des poids optimaux w_0 sujet à $\|w_0\|^2$ est minimum pour pouvoir maximiser la marge $m = 2\|w_0\|^{-1}$ (Scholkopf et al., 1999) et $y_i.(w_0.x_i + b_0) \geq 1$ pour $i=1,2,..., L$. Les machines à supports de vecteurs sont les exemples d'apprentissages x_i qui satisfont l'égalité suivante:

$$y_i.(w_0.x_i + b_0) = 1 \quad (2.45)$$

L'équation précédente définit deux hyperplans; l'un contient les vecteurs à support de la classe -1 et l'autre contient les vecteurs à support de la classe +1. La distance entre les deux hyperplans est maximisée lorsque la norme du vecteur de poids $\|w_0\|^2$ est minimale. La minimisation est obtenue par la maximisation de la fonction suivante :

$$W(\alpha) = \sum_{i=1}^L \alpha_i - 0.5 \sum_{i=1}^L \sum_{j=1}^L \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \quad (2.46)$$

Sujet à,

$$0 \leq \alpha_i \quad \text{et} \quad \sum_{i=1}^L \alpha_i y_i = 0 \quad (2.47)$$

Où α est le multiplicateur de Lagrange. Si $\alpha_i > 0$ alors x_i correspond à un support de vecteur. La classification d'un vecteur d'entrée inconnu est obtenue selon le calcul suivant :

$$F(x) = \text{sign}\{w_0.x + b_0\} \quad (2.48)$$

Où,

$$w_0 = \sum_{i=1}^L \alpha_i \cdot y_i \cdot x_i \quad (2.49)$$

Finalement, la classification est obtenue en effectuant le produit scalaire du vecteur des poids optimaux w_0 et les vecteur d'entrée x .

La Figure 2.8 illustre un exemple du problème de classification binaire et linéairement séparable. Il y a deux classes dénotées par $y = 1$ (des points blancs) et $y = -1$ (points noirs). $H1$ et $H2$ sont les hyper-plans parallèles qui passent par les points qui sont les plus proches à la frontière représentée par la ligne hachurée. La distance entre $H1$ et $H2$ est la marge. Comme mentionné auparavant, la solution du problème de minimisation (réduisant au minimum la norme de w) permet de déterminer l'hyper-plan optimal de décision qui maximise la marge entre les deux classes $y = 1$ et $y = -1$. Dans ce cas-ci, classes peuvent être linéairement séparées par une fonction noyau de type linéaire.

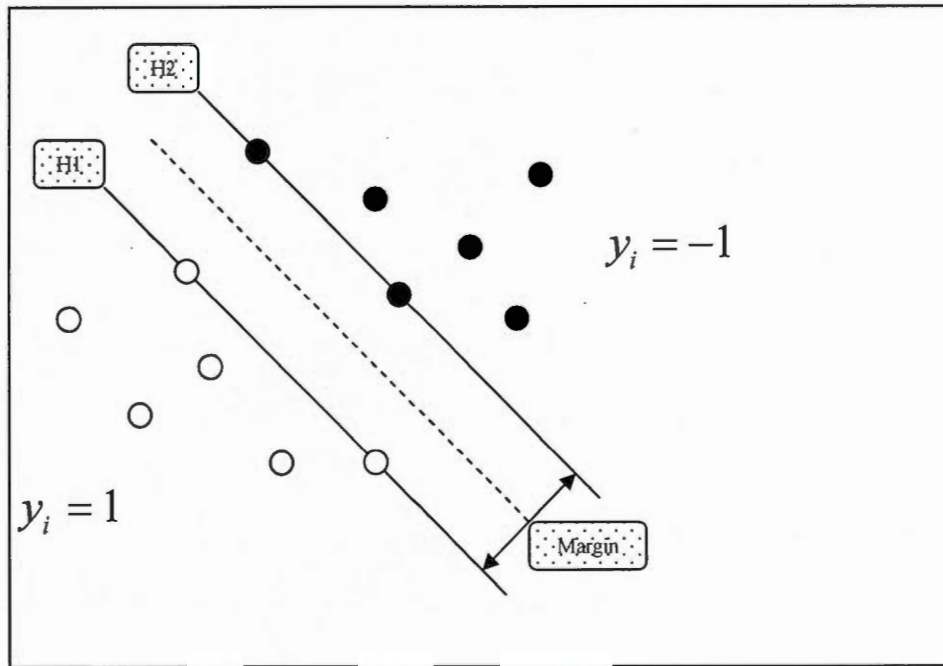


Figure 2.8 Problème de classification binaire et linéairement séparable

2.5.2 Le cas des données non linéairement séparables et SVM non linéaire

Dans le cas des données linéairement séparables bruitées, des variables $(\xi_1, \xi_2, \dots, \xi_L)$ qui indique si l'entrée est classée du bon côté est introduite. En particulier on a $\xi_i \geq 0$ sujet à $y_i \cdot (w_0 \cdot x_i + b_0) \geq 1 - \xi_i$, $i=1, 2, \dots, L$. Alors, $\sum_{i=1}^L \xi_i$ est la bande supérieure des erreurs de classifications. L'hyperplan optimal est la solution au problème d'optimisation suivant :

$$\text{Minimiser } \frac{1}{2} w \cdot w + C \cdot \sum_{i=1}^L \xi_i \quad (2.50)$$

Sujet à :

$$y_i \cdot (w \cdot x_i + b_0) \geq 1 - \xi_i \quad \text{et} \quad \xi_i \geq 0, \quad i=1, \dots, L \quad (2.51)$$

Où C est un paramètre de régularisation qui contrôle le compromis entre la capacité d'apprentissage et l'ensemble des erreurs de l'apprentissage. Par conséquent, une grande valeur de C correspond à une grande pénalité pour les erreurs.

Pour obtenir le SVM non-linéaire dans le cas des données non-linéairement séparables, les entrées sont projetées dans un espace de plus grande dimension en utilisant une fonction de projection K prédéterminée (Cortes et Vapnik, 1995; Scholkopf et al., 1999; Scholkopf et al., 2000). De ce fait, l'hyperplan non linéaire et optimal est construit dans ce nouvel espace. Selon le théorème de Mercer (Scholkopf et Smola, 2002), il existe une fonction de projection K tel que :

$$W(\alpha) = \sum_{i=1}^L \alpha_i - 0.5 \sum_{i=1}^L \sum_{j=1}^L \alpha_i \alpha_j K(x_i, x_j) y_i y_j \quad (2.52)$$

Sujet à,

$$0 \leq \alpha_i \leq C \quad \text{et} \quad \sum_{i=1}^L \alpha_i y_i = 0 \quad (2.53)$$

La fonction de décision (classification) devient :

$$F(x) = \text{sign} \left(\sum_{i=1}^L \alpha_i y_i K(x, x_i) + b_0 \right) \quad (2.54)$$

Les fonctions noyau ou Kernel les plus utilisées sont le noyau de type polynomial et le noyau de type gaussien définis respectivement comme suit :

$$K(x, x_i) = ((x_i \cdot x) + 1)^d \quad (2.55)$$

$$K(x, x_i) = \exp \left(\frac{-\|x - x_i\|^2}{\sigma^2} \right) \quad (2.56)$$

Où d est l'ordre de la fonction noyau polynomiale et δ est l'écart type de la gaussienne.

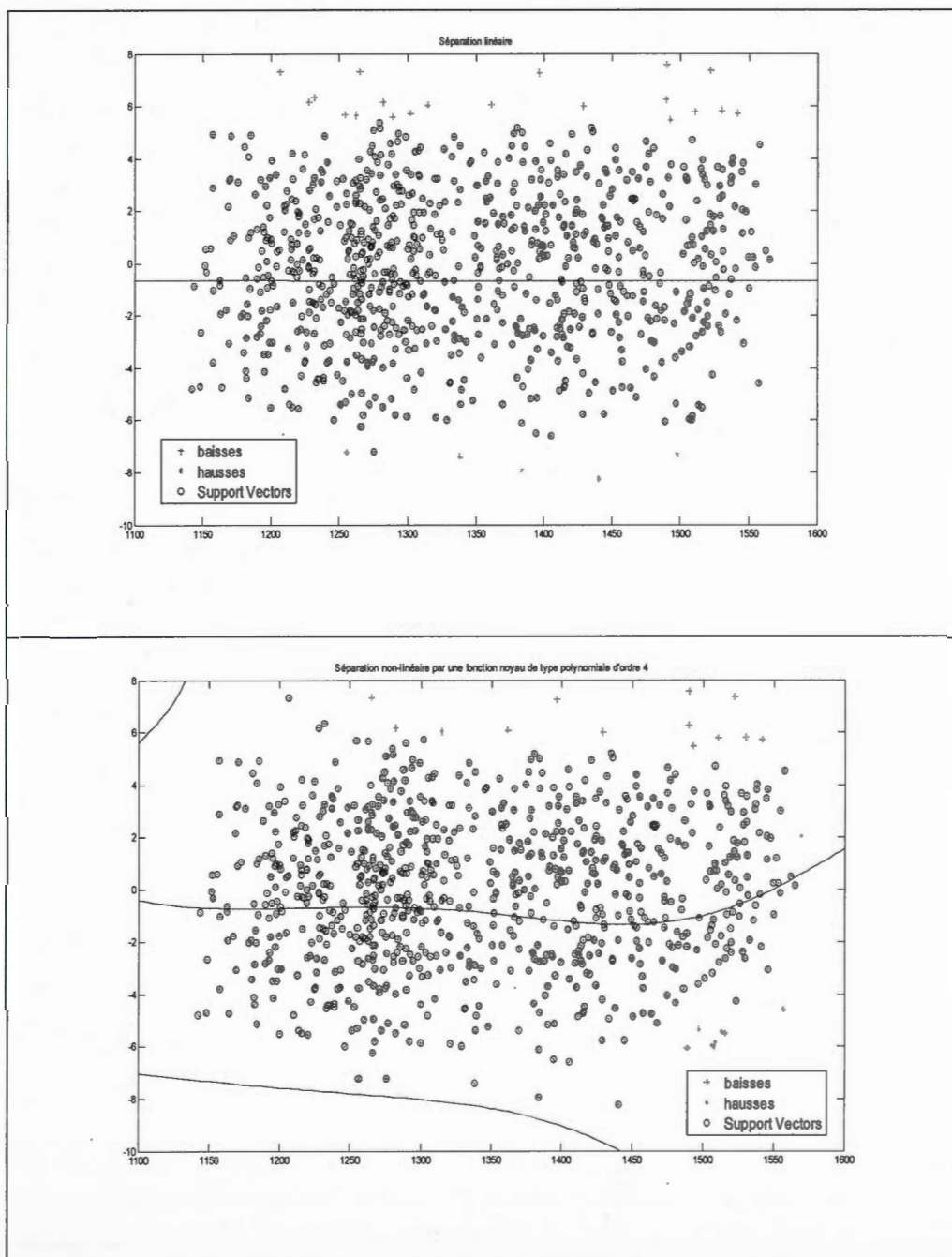


Figure 2.9 Séparation des données : fonction noyau linéaire versus polynomiale

Comme mentionné auparavant, dans le cas des données non-linéairement séparables, une fonction noyau non-linéaire devrait être utilisée pour séparer les données. La Figure 2.9 montre l'hyper-plan optimal utilisant une fonction noyau linéaire (en haut) et l'hyper-plan optimal utilisant une fonction noyau non-linéaire de type polynomial d'ordre quatre (en bas). Dans les deux cas, les séries temporelles du rendement (variation) du marché boursier (S&P500) ont été décomposés par l'analyse en paquet d'ondelette. Les informations obtenues à savoir les coefficients d'approximation et les coefficients de détails constituent les entrées aux SVM pour séparer les tendances haussières des tendances baissières.

Dans le cadre de notre thèse, la fonction noyau polynomiale d'ordre deux, trois, et quatre est retenue pour effectuer les simulations parce qu'elle est une fonction noyau globale et donc permet à tous les points des données d'avoir une influence sur la valeur du noyau (Müller et al., 2001; Smit et Jordaan, 2002). Il n'a pas été nécessaire de considérer des ordres supérieurs à quatre car nous avons constaté que le problème d'optimisation devient plus complexe, prend plus de temps et de mémoire et ne converge pas souvent sans pour autant améliorer les résultats.

La fonction noyau de type gaussien n'est pas retenue car elle est locale et donc seulement les points des données qui sont proches entre elles ont une influence sur la valeur du noyau (Müller et al., 2001; Smit et Jordaan, 2002). Le paramètre C qui contrôle le compromis entre la capacité d'apprentissage et l'ensemble des erreurs de l'apprentissage a été varié pour prendre les valeurs 5, 10, et 15. Seuls les résultats de la valeur 10 seront présentés car ils fournissaient en général les meilleurs résultats. Il faut tout de même noter que les différences étaient très petites.

2.6 Normalisation des données et mesures de performances

Pour éviter des valeurs larges pour les entrées et assurer une meilleur convergences des systèmes prédictifs, toutes les entrées seront normalisées (Cano et al., 2009; Wang et Huang, 2009, Fabricio, 2011) dans l'intervalle $[-1,+1]$ selon la formule suivante :

$$x' = \frac{(2 * x - (Max[x] + Min[x]))}{Max[x] - Min[x]} \quad (2.57)$$

Finalement, l'évaluation de la performance de chaque système est mesurée par le taux de détection, la sensibilité, et la spécificité qui sont données comme suit :

$$Détection = \frac{ECC}{EC} \quad (2.58)$$

$$Sensibilité = \frac{EPCC}{VEP} \quad (2.59)$$

$$Spécificité = \frac{ENCC}{VEN} \quad (2.60)$$

Où ECC est l'échantillon correctement classifié, EC est l'échantillon classifié, EPCC est l'échantillon positif correctement classifié, VEP est le vrai échantillon négatif, ENCC est l'échantillon négatif correctement classifié, et VEN est le vrai échantillon négatif.

Le taux de détection mesure la classification correcte des hausses et des baisses du marché boursier. Par ailleurs, dans notre définition du problème, la sensibilité mesure le taux d'erreur de classification des tendances en baisse; par exemple prévoir une vraie baisse du marché comme étant une hausse éventuelle. La spécificité mesure le taux d'erreur de classification des tendances en hausse; par exemple prévoir une vraie hausse du marché comme étant une baisse éventuelle.

Dans le cadre de cette thèse, nous ne pouvions pas effectuer une validation croisée au sens habituel pour calculer la moyenne et l'écart type du taux de classification parce que nous faisons de la prédiction basée sur une seule série temporelle et donc nous ne pouvons pas utiliser le futur pour prédire le passé. De plus, il est important de préserver la chronologie des événements car elle possède de l'information sur la dynamique des entrées et de la variable à prédire. En somme, il s'agit d'un problème de classification dans le contexte d'analyse de séries chronologiques dont le facteur temps est déterminant de la dynamique des informations qui circulent. D'ailleurs, les études précédentes qui traitent le problème de prédiction boursière dans un cadre de classification de hausse ou de baisse de l'indice boursier n'ont pas adopté la technique de validation croisée (Fernandez-Rodriguez et al., 2000; Harvey et al., 2000; Tay et Cao, 2001; Halliday, 2004; Doesken et al., 2005; Huang et al., 2005; Atsalakis et Valavanis, 2009; Azzini et al., 2010; Cengiz et Herwartz, 2011; Nyberg, 2011; Wang et al., 2011). Néanmoins, nous avons fait une validation croisée à 10 blocs aléatoires et avons calculé la moyenne et l'écart type du taux de classification correct. L'objectif était seulement de rapporter ces résultats à titre indicatif et nous laissons le soin au lecteur d'évaluer leur pertinence. Par ailleurs, il faudrait rappeler que, contrairement aux modèles statistiques linéaires comme la régression (moindre carré ordinaire, moindre carré générale, moindre carré pondérée, méthode des moments généralisés), notre usage des RNA et SVM ne permet pas de faire de l'inférence statistique pour juger de la qualité des résultats.

De ce fait, le taux de détection, la sensibilité, et la spécificité qui sont présentés ci haut sont les mesures de performance les plus appropriées pour évaluer nos résultats.

CHAPITRE III

DONNÉES ET PRÉTRAITEMENTS

Dans ce chapitre, nous allons présenter les données au niveau de la section 3.1 et les prétraitements au niveau de la section 3.2. Noter que les détails techniques sont donnés dans les annexes.

3.1 Les données

Dans notre étude, nous considérons l'indice boursier Standard&Poors 500 (S&P500). Notre choix de cet indice boursier est justifié par le fait qu'il est largement considéré dans la littérature (Choi et al., 1995; Tsai et al., 1998; Ritanjali et Panda, 2007; Yudong et Lenan, 2009; Azzini et al., 2011; Nyberg, 2011) et par son importance économique : il représente les 500 plus grandes capitalisations boursières aux États-Unis d'Amérique. De plus, on peut comparer nos résultats obtenus à ceux de la littérature. La période d'étude s'étale sur la période allant du 21 octobre 2003 au 29 janvier 2008 après prétraitements. En effet, les données doivent être prétraitées afin de garantir un meilleur apprentissage et une convergence efficace. Le prétraitement des données comprend quatre étapes importantes :

a) Déterminer les processus²³ ARMA²⁴ et EGARCH²⁵ de la variation de l'indice boursier (voir plus bas).

²³ Les processus ARMA (ARIMA) et EGARCH sont employés pour modéliser les mouvements du rendement du marché boursier afin d'extraire sa volatilité.

- b) Vérifier la stationnarité de toutes les variables d'entrée.
- c) Effectuer les tests de causalité pour sélectionner les variables qui ont une relation causale avec la variation future de l'indice boursier. .
- d) Normaliser toutes les variables dans l'intervalle $[-1,1]$ au besoin.

Les méthodes statistiques seront employées pour rendre les données stationnaires, modéliser le processus des rendements et extraire sa volatilité, et finalement pour déterminer les variables prédictives. Le prétraitement statistique des données est présenté au niveau de ce chapitre. Les détails techniques sont présentés dans les annexes afin d'alléger la lecture du texte.

En statistiques, le modèle ARMA(p,q) (Box et Jenkins, 1970) est couramment utilisé pour représenter un processus stochastique linéaire et stationnaire. Il exprime la dynamique linéaire du processus à travers deux composantes. La première est dite autorégressive d'ordre p car elle repose sur les p retards les plus significatifs de la variable aléatoire étudiée. La deuxième composante est dite moyenne mobile d'ordre q car elle représente la tendance à long terme du processus en termes des q erreurs retardées les plus significatives. Le processus ARMA(p,q) sert donc à modéliser la moyenne conditionnelle d'une variable aléatoire temporelle, par exemple le rendement boursier $R(t)$ définit plus tôt.

²⁴ Puisque la séries temporelle des rendements (variations du prix $P(t)$ de l'indice boursier S&P500 exprimé en $R(t)$) est stationnaire (Voir Annexe C) alors le processus ARMA(p,q) que suit $R(t)$ équivaut à un processus ARIMA(p,i,q) où $i=1$. En d'autres termes, le prix de l'indice boursier est intégré d'ordre un. En somme, nous modélisons $R(t)$ par un processus ARIMA(p,0,q). Les paramètres p, q, et i sont respectivement l'ordre de la composante autorégressive, l'ordre de la composante moyenne mobile, et l'ordre d'intégration de la série.

²⁵ La volatilité du rendement boursier mesurée par la série EGARCH qui fera partie des mesures de sentiment comme on le verra plus loin.

Le modèle EGARCH(m,n) est employé pour capturer la variance conditionnelle des erreurs qui découlent du processus ARMA(p,q) en termes de m retards et de n retards de résidus, tenant compte des phénomènes d'asymétries dans les erreurs et considérant que la variance conditionnelle n'est sans doute pas la même selon que l'erreur précédente est positive ou négative. Dans notre cas, il sert à modéliser la volatilité estimée de la variable $R(t)$. Par ailleurs, le processus EGARCH assure aussi la positivité de la variance conditionnelle estimée.

Le modèle ARMA constitue une étape importante avant d'utiliser le modèle EGARCH (Nelson, 1991). En effet, la qualité de l'estimation de la volatilité par EGARCH dépend d'abord de la qualité du modèle ARMA employé pour modéliser la moyenne conditionnelle. Pour choisir le modèle ARMA qui est le mieux ajusté à la série de données et le moins complexe, le critère AIC (Akaike Information Criterion) et le critère BIC (Bayesian Information Criterion) sont généralement utilisés (Carmona, 2003). Ces deux critères permettent de sélectionner le modèle dont la vraisemblance est grande tout en la pénalisant par la complexité du modèle.

La description générale de l'implémentation des approches ARMA et EGARCH pour la modélisation de la moyenne conditionnelle et de la variance conditionnelle des rendements $R(t)$ est donnée plus loin au niveau de ce chapitre. Les détails techniques sont donnés au niveau de l'annexe II.

Les variables de prédictions sont de trois types: variables macroéconomiques, indicateurs techniques et mesures du sentiment. Toutes les données sont à fréquence quotidienne. La série temporelle de l'indice boursier S&P500 est donnée dans la Figure 3.1.

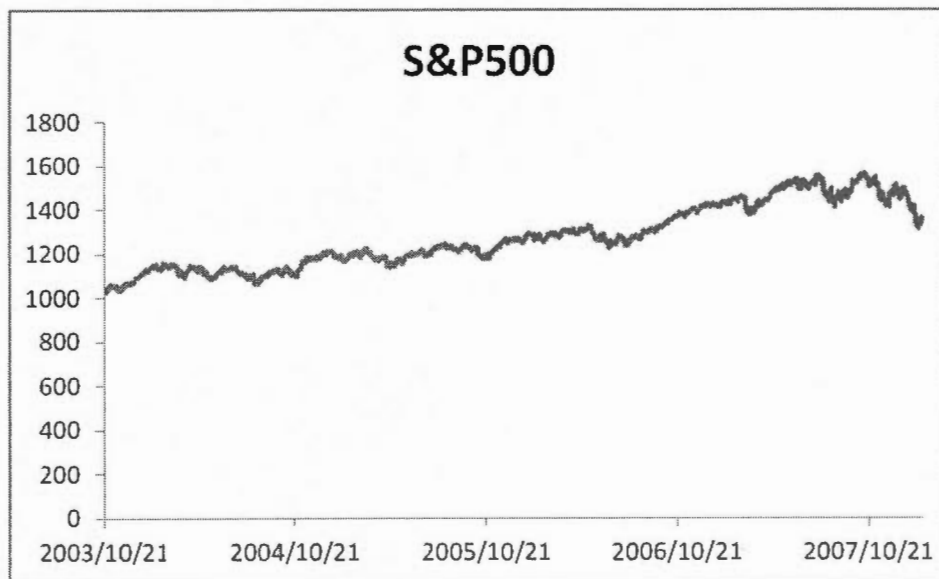


Figure 3.1 Série temporelle du S&P500

Rappelons que pour prédire la tendance à la hausse ou à la baisse de l'indice boursier S&P500, on devrait approximer la fonction suivante :

$$y_t = f(x_t) \quad (3.1)$$

Où t est l'indice de temps, x est la variable d'entrée ou variable prédictive, et y est la variable de sortie définie comme suit :

$$y_t = +1 \quad \text{si} \quad R_t > 0 \quad (3.2)$$

$$y_t = -1 \quad \text{si} \quad R_t < 0 \quad (3.3)$$

Où, R est le taux de variation de l'indice boursier I .

$$R_t = \log\left(\frac{I_t}{I_{t-1}}\right) \quad (3.4)$$

Noter qu'il est nécessaire d'obtenir la série R pour pouvoir effectuer les tests de causalité de Granger (1969) et aussi pour pouvoir modéliser et extraire la volatilité de l'indice boursier S&P500 qui sera parmi les mesures de sentiments comme on le verra ultérieurement au niveau de ce chapitre.

La série temporelle ($R(t)$) de l'indice boursier S&P500 est donnée dans la Figure 3.2. L'histogramme et les statistiques descriptives de $R(t)$ sont donnés dans la Figure 3.3. Nous avons considéré toutes les variables macroéconomiques et monétaires fréquemment utilisées dans la littérature et aussi considérées par les communautés des analystes financiers et des économistes²⁶. Elles sont extraites de la base de données appelée FRED II Database²⁷ de la banque fédérale de Saint-Louis au Missouri. La liste²⁸ des variables macroéconomiques est donnée dans le Tableau 3.1. Toutes les séries temporelles des variables macroéconomiques sont montrées dans la Figure 3.4.

²⁶ Hardouvelis (1987), Schwert (1990), Flannery et Protopapadakis (2002), Yiwen et al. (2000), Wu et al. (2001), Egeli et al. (2003), Thawornwong et Enke (2004), Chen et al. (2003), Huang et al. (2005).

²⁷ <http://research.stlouisfed.org/fred2/>

²⁸ La liste des variables macroéconomiques n'inclut pas le produit intérieur brut (PIB) comme variable prédictive puisque nous utilisons des variables observées quotidiennement pour prédire la tendance du marché boursier. Par contre, les séries temporelles du PIB sont observées trimestriellement. De ce fait, pour utiliser les données sur le PIB il faudrait effectuer une extrapolation qui fait augmenter le risque d'imprécision en estimant des données quotidiennes à partir des données trimestrielles. De plus, les estimés erronés engendrent un certain bruit qui affecte négativement l'apprentissage du RNA. En effet, l'extrapolation n'est pas recommandée pour les raisons suivantes : (i) pour une variable donnée, le taux de croissance trimestriel est plus grand que le taux de croissance quotidien, (ii) il y a une possibilité que le régime et la tendance quotidienne ne correspondent pas au régime et à la tendance trimestrielle, (iii) les conditions économiques et financières quotidiennes qui sont responsables des variations des prix sont différentes des conditions trimestrielles, (iv) il est difficile de choisir le modèle le plus approprié pour réaliser l'extrapolation et obtenir les meilleurs estimés quotidiens à partir des données trimestrielles, et (v) les données quotidiennes sur les taux de changes et indicateurs des exportations américaines et les entreprises cotées AAA représentent de meilleurs indicateurs sur les conditions macroéconomiques que les estimés du PIB obtenus par extrapolation à partir des données trimestrielles.

D'autre part, nous utilisons également les indicateurs techniques les plus fréquemment utilisés dans la littérature²⁹.

La liste des indicateurs techniques est donnée dans le Tableau 3.2 et leurs séries temporelles sont présentées dans la Figure 3.5 et Figure 3.6. Notons que la variable à prédire et toutes les variables prédictives sont mesurées à la même fréquence (quotidienne) y compris les mesures de sentiment que nous allons présenter ultérieurement. Rappelons qu'à l'instar de la série temporelle de l'indice boursier du S&P500, toutes les variables prédictives couvrent la période allant du 21 octobre 2003 au 29 janvier 2008.

Finalement, 80% des données sont consacrée à l'entraînement du RNA et du SVM, et 20% sont consacrée à la phase des tests (simulations). Notons aussi - comme nous allons le voir ultérieurement- que la taille de l'ensemble des variables économiques, techniques et de sentiment sera respectivement réduite après avoir appliqué le test de Granger (1969) pour déterminer les variables qui statistiquement causent un changement dans le marché boursier.

²⁹ Lendasse et al., (2000), Motiwalla et Wahab (2000), Leigh et Purvis (2002), Bautista (2001), Dong et al. (2003), Armano et al. (2004), Jaruszewicz et Mandziuk (2004), Baek et Cho (2002), Tan et al. (2011), Lam (2001), Dong et Zhou (2002), Dourra et Siy (2002).

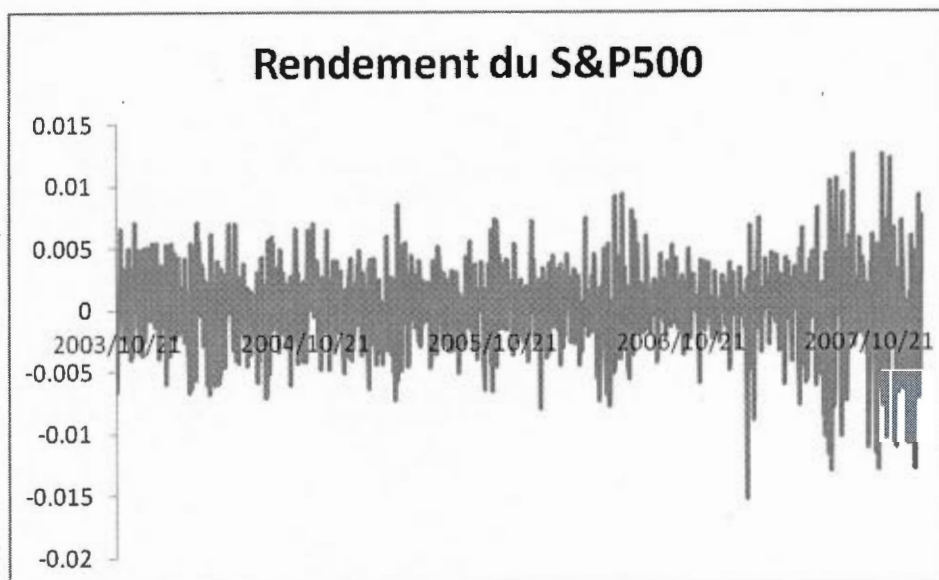


Figure 3.2 Série temporelle $R(t)$ du S&P500

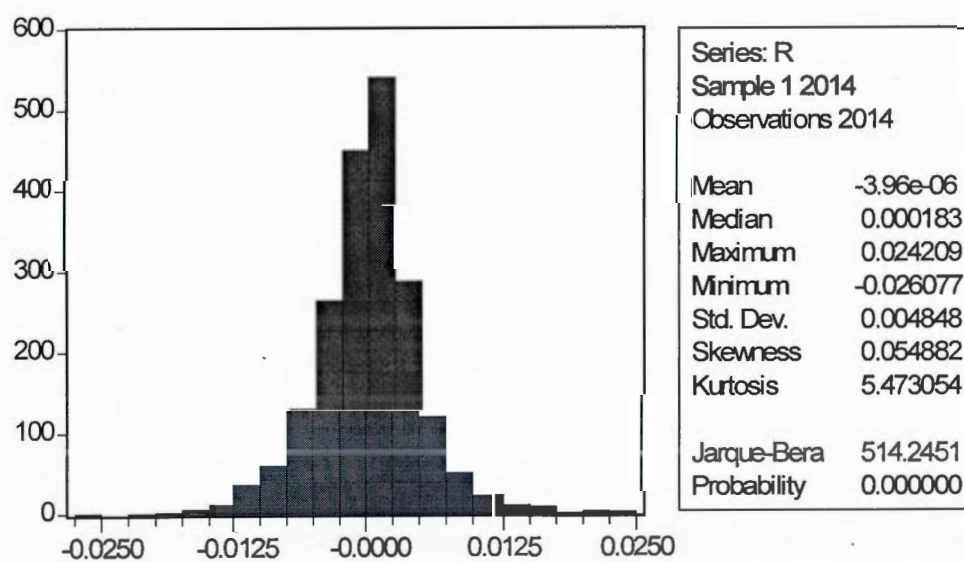


Figure 3.3 Histogramme et statistiques descriptives de $R(t)$

Tableau 3.1

Liste des variables économiques

Type de variables	Description
<i>Variables financières</i>	
DAAA	Moody's Seasoned Aaa Corporate Bond Yield
DBAA	Moody's Seasoned Baa Corporate Bond Yield
DTB3	3-Month Treasury Bill: Secondary Market Rate
DTB6	6-Month Treasury Bill: Secondary Market Rate
<i>Variables monétaires</i>	
DFEDTAR	Federal Funds Target Rate
DFF	Effective Federal Funds Rate
<i>Taux de changes</i>	
DEXCAUS	Canada / U.S. Foreign Exchange Rate
DEXJPUS	Japan / U.S. Foreign Exchange Rate
DEXSZUS	Switzerland / U.S. Foreign Exchange Rate
DEXUSEU	U.S. / Euro Foreign Exchange Rate
DEXUSUK	U.S. / U.K Foreign Exchange Rate
DTWEXB	Trade Weighted Exchange Index: Broad
DTWEXM	Trade Weighted Exchange Index: Major Currencies

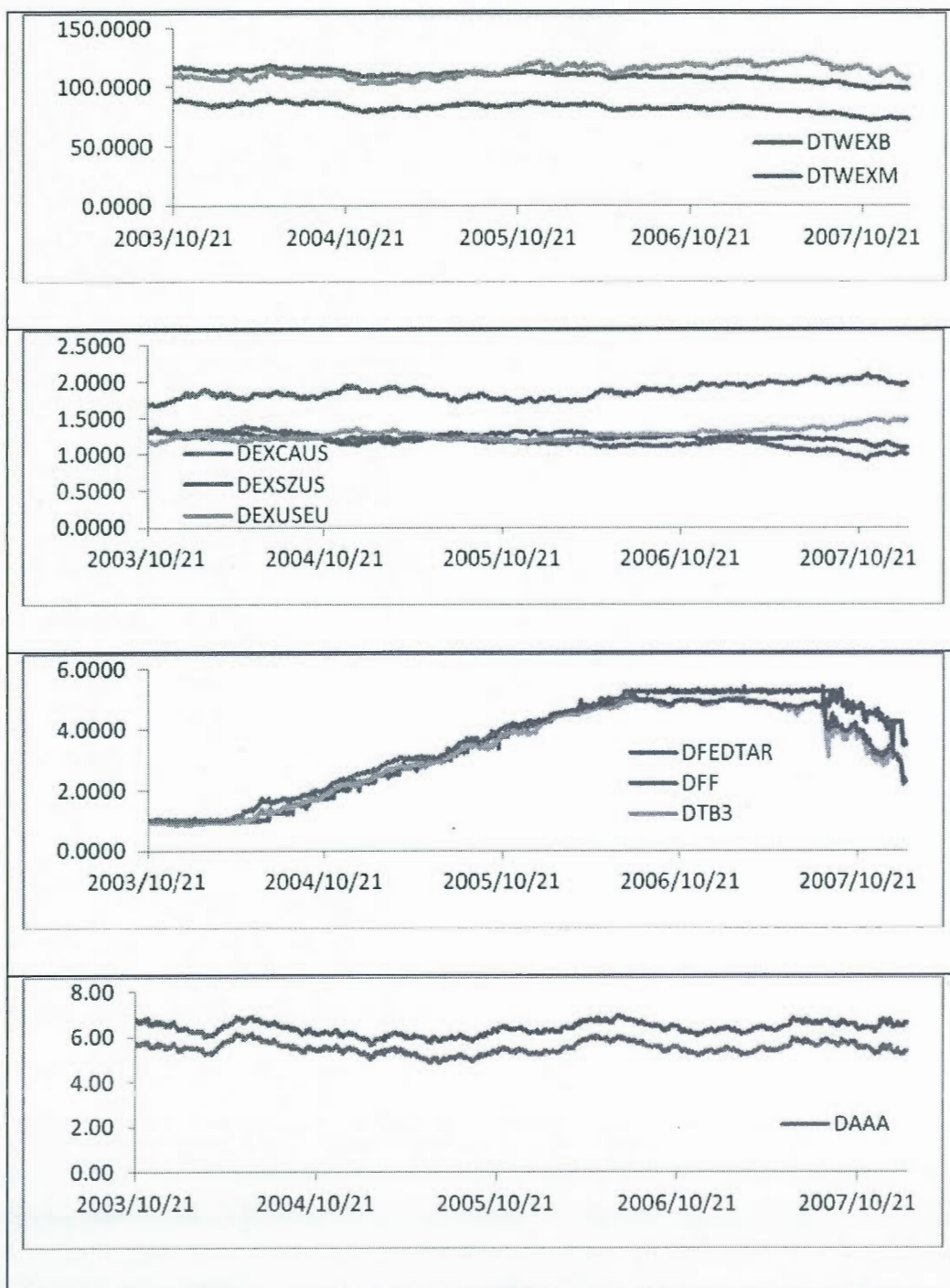


Figure 3.4 Séries temporelles des variables économiques

Tableau 3.2

Liste des indicateurs techniques

Indicators	Description
MA(t)	Moyenne mobile
X1(t)	$[X(t) - PPB(t)] / [(PPH(t) - PPB(t))]$
X2(t)	$[X(t) - MA(t)] / MA(t)$
X3(t)	$X(t) - \min[X(t), \dots, X(t-5)]$
X4(t)	$X(t) - \max[X(t), \dots, X(t-5)]$
MPM(t)	$[(PPH(t) - PPB(t)) - (PPH(t-2) - PPB(t-2))]/2$
BR(t)	$\text{volume}(t) / [(PPH(t) - PPB(t))/2]$
EMV(t)	$MPM(t) / BR(t)$
MOM(t)	$PC(t) - PC(t-6)$
ROC(t)	$PC(t) / PC(t-6)$
DIS(t)	$[PC(t) / MA(t-6)] * 100$
STOD(t)	$[PC(t) - PPB(t, t-6)] / [PPH(t, t-6) - PPB(t, t-6)]$

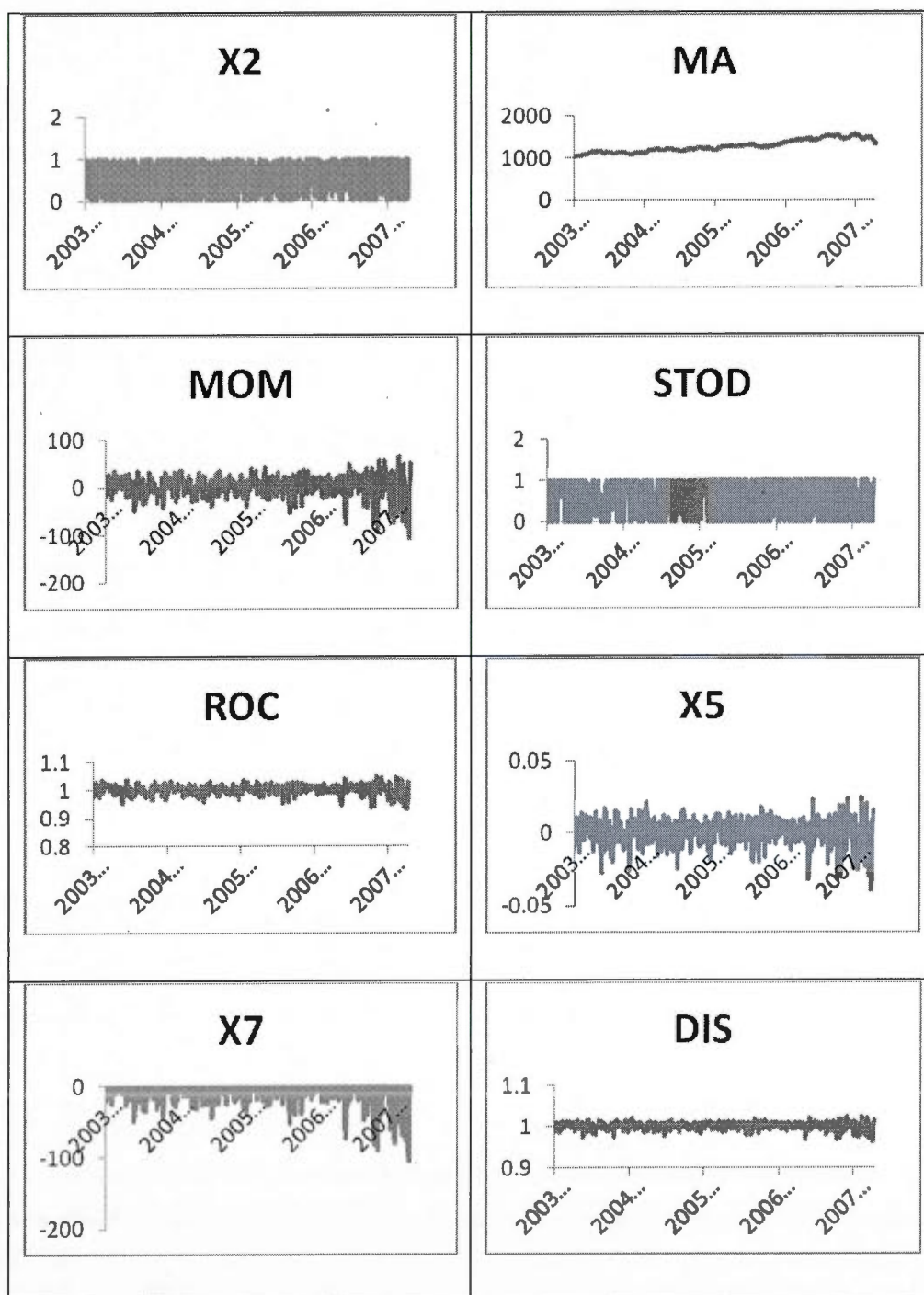


Figure 3.5 Séries temporelles des indicateurs techniques

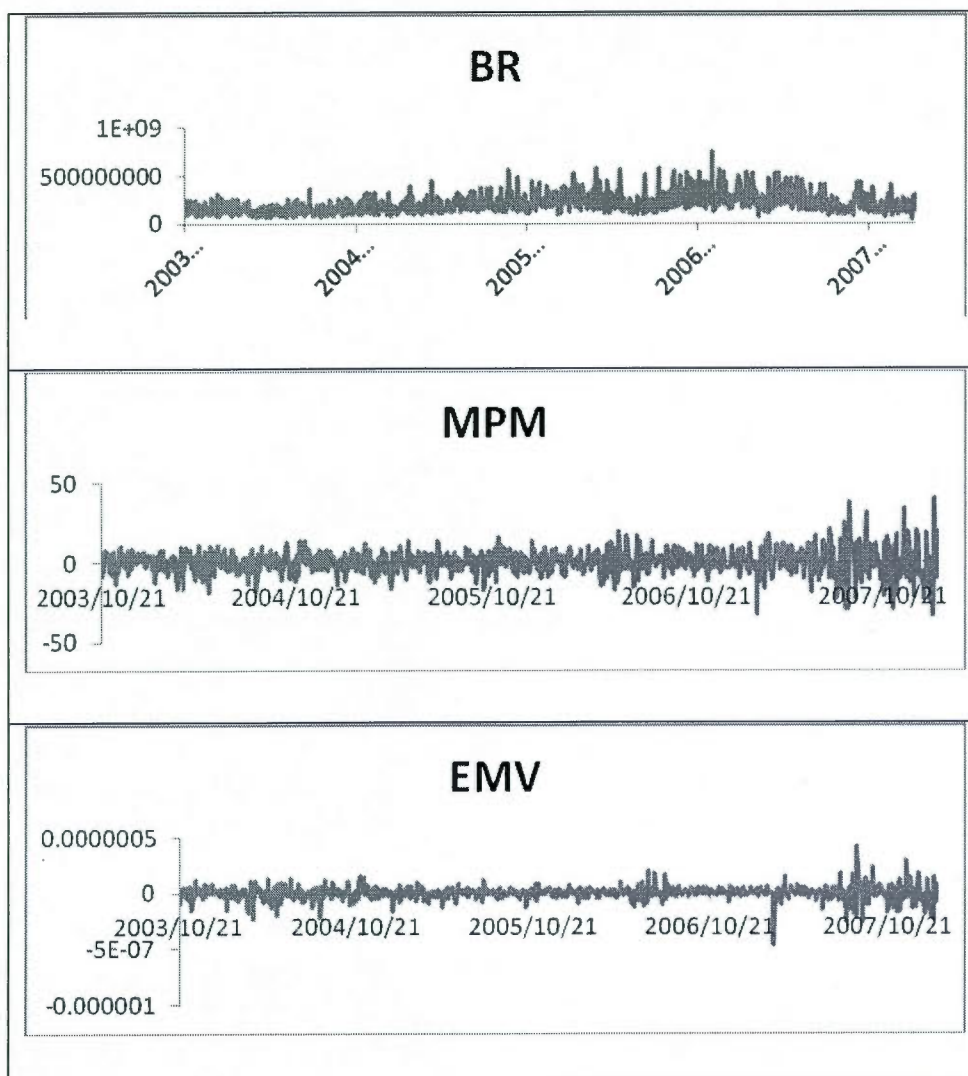


Figure 3.6 Séries temporelles des indicateurs techniques BR, MPM, et EMV

Dans notre thèse, cinq mesures³⁰ du sentiment de l'investisseur les plus utilisées en finance comportementale sont employées pour prédire la tendance à la hausse ou à la baisse du marché S&P500. La première mesure est le " Volatility Index" (VIX) du Chicago Board Options Exchange (Whaley, 2000; Bandopadhyaya et Jones 2008; Ben-Rephael et al., 2010). Le VIX est un estimé de la volatilité des options sur l'indice S&P500. La deuxième mesure est le total du volume des transactions du ratio put-call (Put Call Volume Ration) (Yang et Wu, 2011; Simon et Wiggins, 2001). Ces deux mesures sont considérées comme des indices de la peur (fear indices) : un niveau haut de l'indice indique que le sentiment du marché est à la baisse (Simon et Wiggins, 2001). La troisième mesure du sentiment est le State Street's Investor Confidence Index (ICI) (Warther, 1995; Edwards et Zhang, 1998; Fant, 1999; Ben-Rephael et al., 2010). Cet indice mesure l'attitude de l'investisseur face au risque. Les séries temporelles des indices VIX, PutCall, et l'ICI sont montrés dans la Figure 3.7.

En somme, 13 variables économiques, 12 indicateurs techniques, 5 mesures de sentiments, une série d'approximation et une série de détails pour chaque type de transformée en ondelette (AOD, APO), et la série de tendance du marché boursier S&P constituent l'ensemble des séries temporelles utilisées au niveau de cette thèse : soit 34 séries. Donc, l'ensemble des données utilisées est 34680 ($51 \times 20 \times 34$).

³⁰ Il faudrait noter que la littérature propose des mesures de sentiments qui sont subjectives et non pas objectives. En effet, il est difficile de l'approximer. Le lecteur peut consulter la littérature sur le sujet.

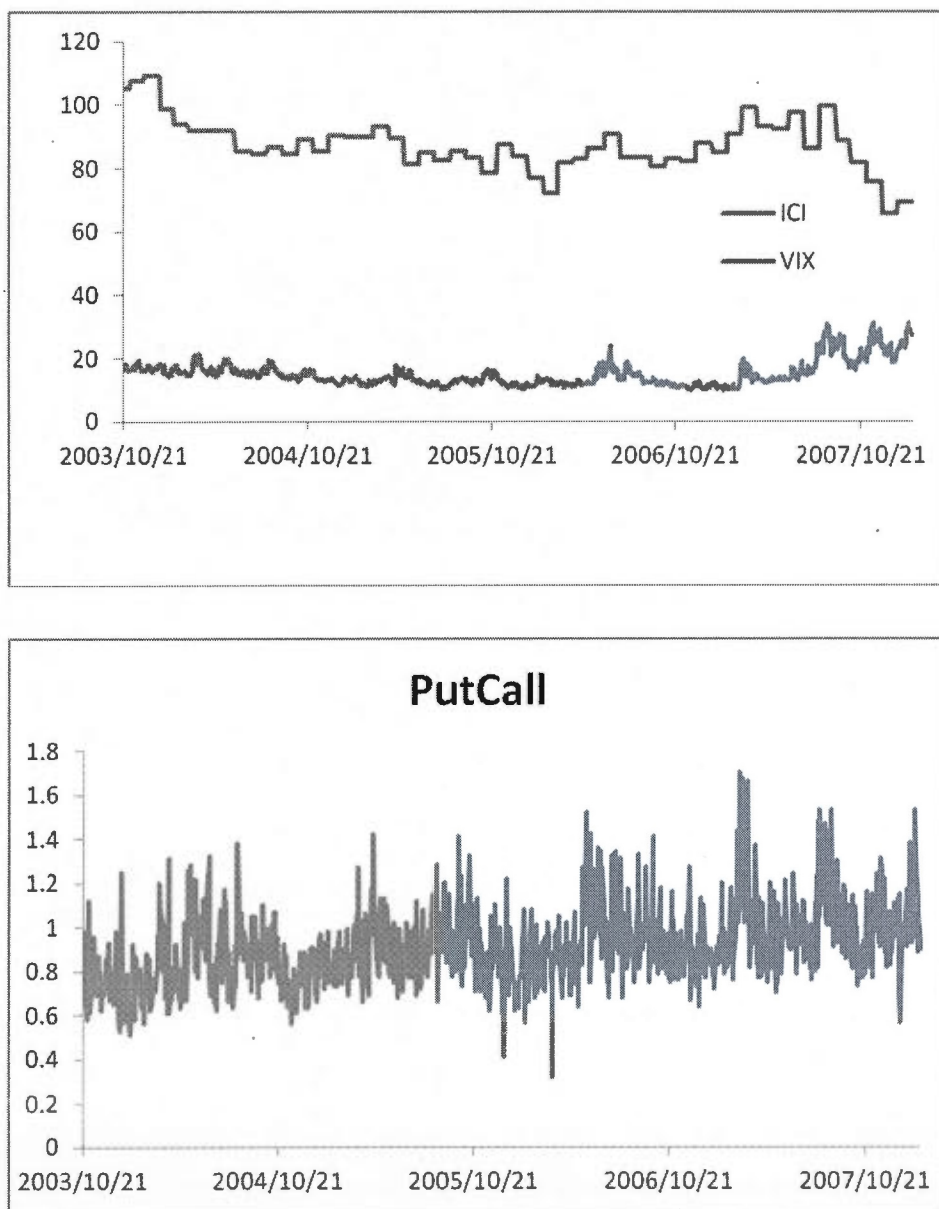


Figure 3.7 Séries temporelles des mesures de sentiment : ICI, VIX, Put-Call

Finalement, le logarithme des séries du volume des transactions et les séries de la volatilité³¹ des variations ($R(t)$) de l'indice boursier sont respectivement quatrième et cinquième mesures du sentiment considérées dans notre thèse (Barber et al.; 2009, Yang et Wu, 2011; Yu et Yuan, 2011 ; Lawrence et Makridakis, 1989 ; Yu et Yuan, 2011 ; Wang et al., 2006; Schmeling, 2007; Verma et Verma, 2007). Leurs séries temporelles sont montrées dans les Figures 3.8 et 3.9.

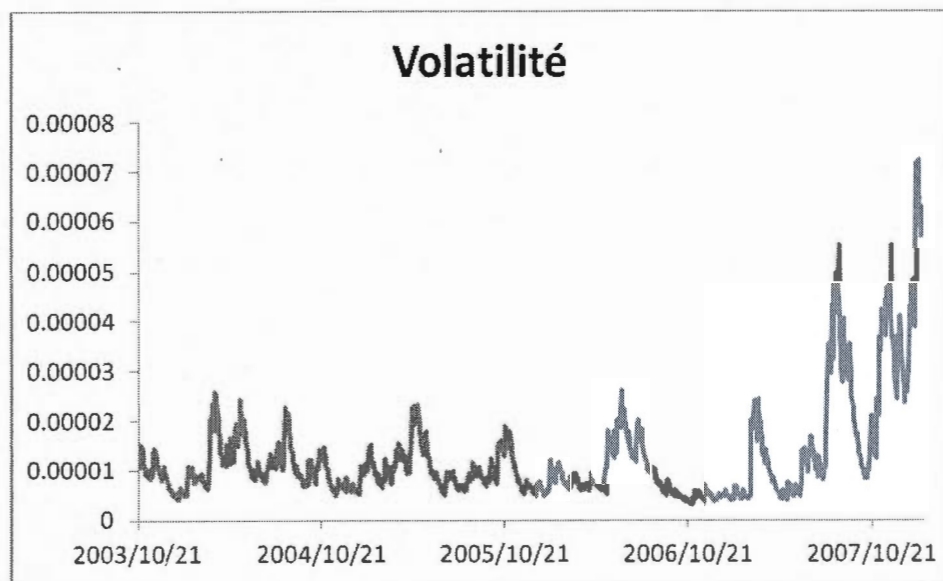


Figure 3.8 Série temporelle de la volatilité du S&P500

³¹Les séries temporelles de la volatilité ont été estimées par processus EGARCH. Cette technique est détaillée en annexe E.

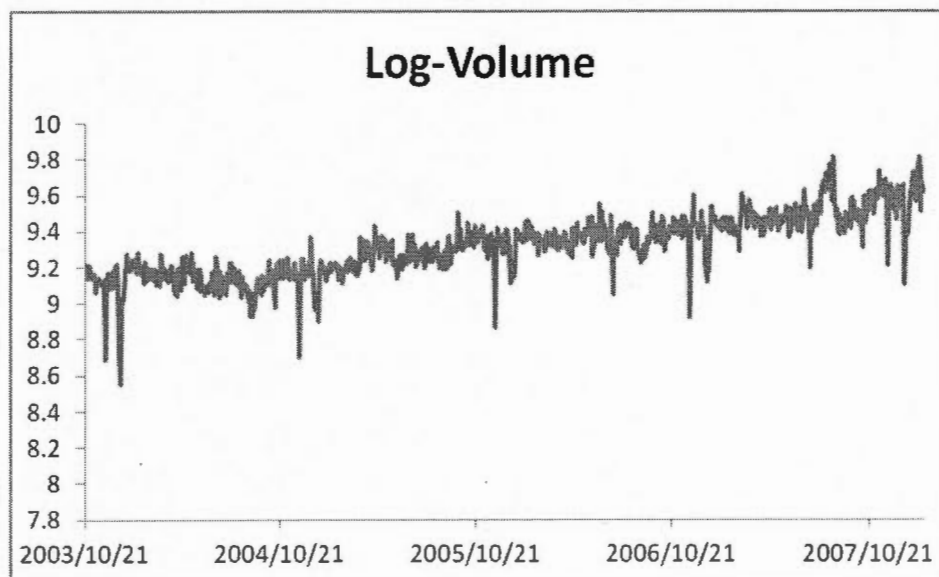


Figure 3.9 Série temporelle du volume de transactions

Il faudrait noter qu'il existe d'autres mesures de sentiments de l'investisseur qui ont été proposés dans la littérature. On peut citer par exemple la décote sur les fonds à capital fermé et le rachat net des fonds mutuels (Neal et Wheatley, 1998). En effet, selon Wiesenberger (1946) l'investisseur est pessimiste au plus haut niveau lorsque la décote sur les fonds à capital fermé est très large et vice versa. De son côté, Malkiel (1977) suggère que le rachat net des fonds mutuels reflète le sentiment général de l'investisseur. Au niveau empirique, Neal et Wheatley (1998) ont conclu que la décote sur les fonds à capital fermé et le rachat net des fonds mutuels la différence de rendement entre les portefeuilles (fonds mutuels) composés d'actions de compagnies à grandes capitalisations et les portefeuilles composés d'actions de compagnies à petites capitalisations.

Dans le cadre de notre thèse, la décote sur les fonds à capital fermé et le rachat net des fonds mutuels (Neal et Wheatley, 1998) ne sont pas employées pour prédire la tendance future de l'indice boursier S&P500 pour les raisons suivantes. Premièrement, les données qui représentent ces mesures de sentiment sont collectées

à partir d'enquêtes sur des portefeuilles hypothétiques ou des fonds mutuels composés d'actions (Neal et Wheatley, 1998). De ce fait, ces mesures de sentiments sont plus adaptées pour prédire les portefeuilles ou les fonds mutuels en question que de prédire un indice boursier qui reflète généralement le marché d'actions. Dans notre thèse, nous modélisons le marché S&P500 pour le prédire et non pas pour prédire des portefeuilles.

Deuxièmement, ces données sont disponibles à une fréquence annuelle contrairement à nos données qui sont de fréquence quotidienne. De ce fait, nos mesures de sentiment sont capables de mieux nous renseigner sur la tendance journalière du marché boursier. En d'autres mots, elles sont plus adaptées à notre problème et à nos données. Quatrièmement, les données pour mesurer le sentiment de l'investisseur et qui sont utilisées dans (Neal et Wheatley, 1998) ne sont pas des données de nature publique ce qui y réduit l'accès. Par contre nos mesures sont publiquement disponibles à n'importe quel investisseur qui voudrait les utiliser. Cinquièmement, Neal et Wheatley (1998) ont conclu qu'on ne peut pas statistiquement conclure que la décote sur les fonds à capital fermé et le rachat net des fonds mutuels affectent significativement les rendements des portefeuilles étudiés. En d'autres mots, la taille des estimateurs obtenus par régression linéaire montrent un effet économique important mais ce dernier n'est pas statistiquement significatif à cause de la taille des erreurs. De ce fait, les résultats obtenus par Neal et Wheatley (1998) restent mitigés. Et sixièmement, les études récentes utilisent de plus en plus les mesures de sentiments que nous adoptons dans notre thèse (Wang et al., 2006; Schmeling, 2007; Verma et Verma, 2007; Bandopadhyaya et Jones 2008; Barber et al., 2009; Ben-Rephael et al., 2010; Yang et Wu, 2011; Yu et Yuan, 2011 ; Yu et Yuan, 2011).

3.2 Processus ARMA et EGARCH

Pour extraire la volatilité ($V(t)$) d'une série temporelle on doit modéliser en premier lieu sa moyenne par le processus ARMA, et en deuxième lieu sa variance conditionnelle par le processus EGARCH (voir annexe II). Par conséquent, il faut bien identifier le bon processus ARMA et également le bon processus EGARCH pour extraire la bonne série de volatilité. Nous appelons ce type de prétraitement processus de modélisation statistique de la série temporelle boursière. Il est montré dans la Figure 3.10.

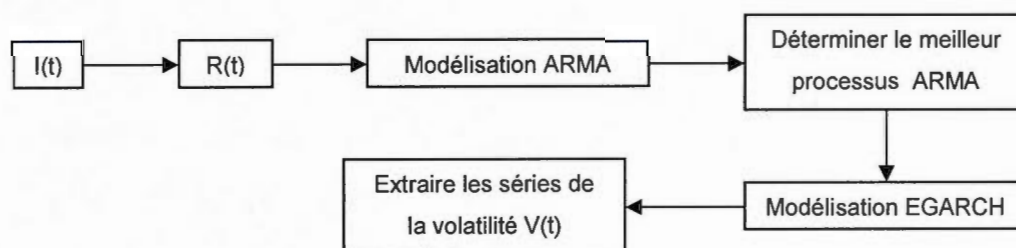


Figure 3.10 Processus de la modélisation statistique de la série temporelle

Ce processus comporte deux principales étapes : la modélisation ARMA et la modélisation EGARCH. La première étape consiste à identifier le bon processus ARMA en minimisant les critères d'informations statistiques Akaike et Schartz. Ensuite, le test d'auto-corrélation des erreurs³² (Breusch et Godfrey, 1981) est appliqué au processus ARMA identifié pour s'assurer que ses erreurs ne sont pas auto corrélées. En d'autres mots, il s'agit de valider le processus ARMA. Si ce n'est pas le cas, on continue à chercher les bons paramètres p et q du processus par les critères d'informations statistiques Akaike et Schartz jusqu'à ce que le bon processus ARMA soit validé par le test de Breusch et Godfrey (1981). Pour rendre l'estimation des paramètres du processus ARMA robuste à l'auto corrélation des résidus, l'estimateur

³²Voir annexe G pour le détail technique et les résultats.

de matrice de covariance de Newey-West (1987) sera employé (voir annexe VI). La Figure 3.11 montre le processus itératif pour la modélisation ARMA³³.

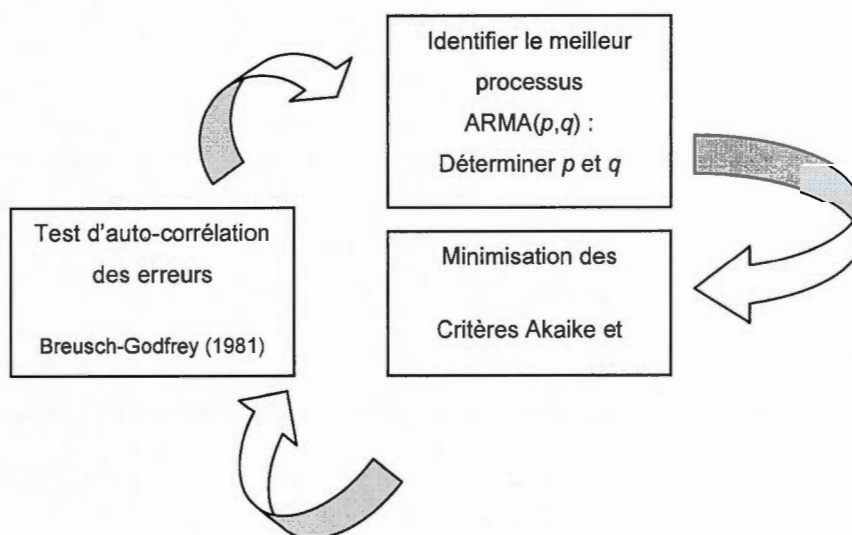


Figure 3.11 La modélisation du processus ARMA(p,q)

Lorsque la première étape qui consiste à déterminer le bon processus ARMA est complétée, alors on passe à la deuxième étape qui consiste à modéliser la variance conditionnelle (la volatilité) de la série temporelle à étudier. En d'autres mots, on cherche à identifier les bon paramètres m et n du processus EGARCH.

À l'instar du processus ARMA, on choisi les paramètres m et n qui minimisent les critères d'informations statistiques Akaike et Schartz. Ensuite, on applique le test d'Engle (1982) pour valider le processus EGARCH identifié. Si ce n'est pas le cas, on recommence la deuxième étape en cherchant d'autres valeurs pour m et n . finalement, lorsque le test d'Engle³⁴ (1982) valide le bon processus EGARCH; alors on peut extraire la série de la volatilité qui est justement représentée par le processus

³³Le détail technique est donné en annexe B.

³⁴Voir annexe H pour le détail technique et aussi les résultats.

EGARCH identifié. La Figure 3.12 montre le processus itératif pour la modélisation du processus EGARCH³⁵.

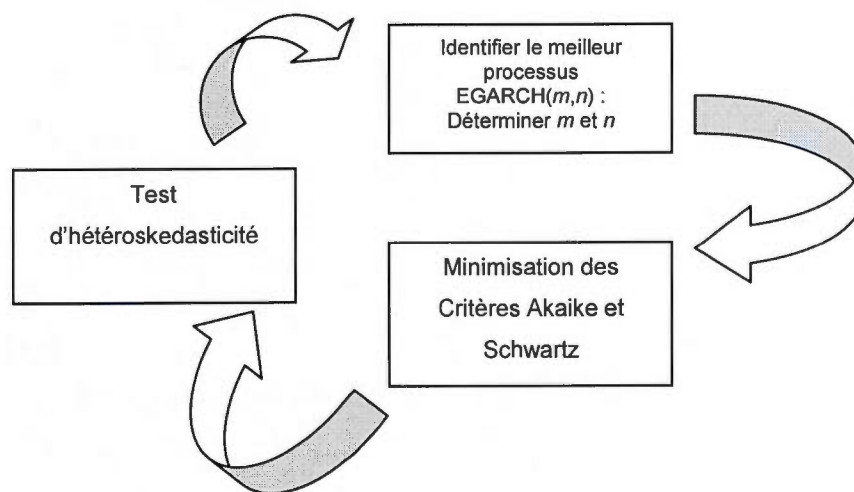


Figure 3.12 La modélisation du processus EGARCH(m,n)

Une fois les séries temporelles des variables prédictives sont obtenues, il va falloir déterminer lesquelles causent au sens de Granger (1969) une variation dans la série temporelle $R(t)$ à prédire. L'objectif est de fournir aux systèmes prédictifs uniquement des entrées qui montrent un lien causal statistiques avec les rendements boursiers. Ce lien est déterminé par le test de Granger³⁶ (1969). De ce fait, on pourrait réduire le bruit qui affecterait le système de classification et qui pourrait être engendré par une variable prédictive non statistiquement reliée au comportement de l'indice boursier. Pour effectuer ce test, il faudrait au préalable s'assurer de la stationnarité statistique des variables en question. Sans cette stationnarité, le test de Granger n'a aucune validité. Le test de stationnarité employé est le test d'Augmented

³⁵Voir annexe H pour les détails techniques.

³⁶Voir annexe D pour la description du test et ses résultats.

Dickey-Fuller (ADF³⁷) proposé par Dickey et Fuller (1979). Dans le cas où une variable n'est pas stationnaire selon le test, il faudra alors effectuer une différence³⁸ de premier ordre ou de deuxième ordre jusqu'à ce que la série soit stationnaire selon ADF. Finalement, c'est la série identifiée stationnaire qui subira le test de Granger en relation avec la série des rendements boursiers $R(t)$.

Le schéma général de ce type de prétraitement préalable au test de causalité de Granger est donné dans la Figure 3.13. En dernier lieu, les séries sélectionnées par le test de Granger vont constituer les entrées finales aux différents systèmes prédictifs. La Figure 3.14 montre un schéma générique qui résume le système prédictif proposé dans notre thèse dont les variables prédictives concernées sont les informations économiques, techniques ou mesures de sentiment.

Au niveau de l'approche par l'usage de l'information historique, le nombre de retards de l'indice boursier S&P500 ($P(t)$) à inclure pour prédire $y(t+1)$ la tendance future de l'indice boursier sera déterminé par l'analyse du comportement de la série ($P(t)$) en examinant sa fonction d'auto corrélation³⁹ partielle, et aussi par la modélisation du processus ARMA(p,q)⁴⁰ de $R(t)$.

Finalement, au niveau de l'approche par analyse par ondelettes; les séries temporelles des basses fréquences obtenues par l'analyse en ondelette discrète (AOD) de l'indice boursier S&P500 ($I(t)$) qui serviront à prédire sa tendance future $y(t)$ sont montrées dans la Figures 3.15. Il s'agit de l'approche standard employée par Wang (2011) qui sera notre modèle d'ondelette de référence. La Figure 3.16 montre les séries temporelles de basses et de hautes fréquences de l'indice boursier S&P500 ($P(t)$)

³⁷Voir annexe C pour la description du test et ses résultats.

³⁸Voir annexe C pour les différences de premier et de deuxième ordre.

³⁹ Voir annexe E pour la description de la fonction et des résultats y afférents.

⁴⁰ Voir annexe C.

obtenues par l'analyse en paquet d'ondelette (APO). Elles constituent les entrées aux classificateurs pour prédire sa tendance future. Il s'agit de l'approche que nous suggérons au niveau de notre thèse.

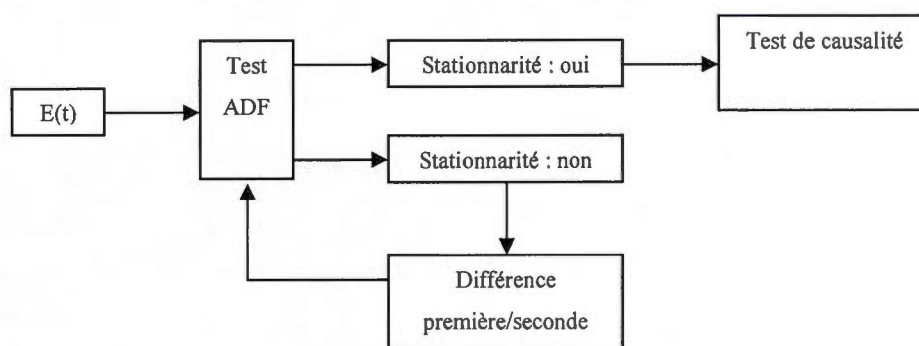


Figure 3.13 Prétraitement préalable au test de causalité de Granger

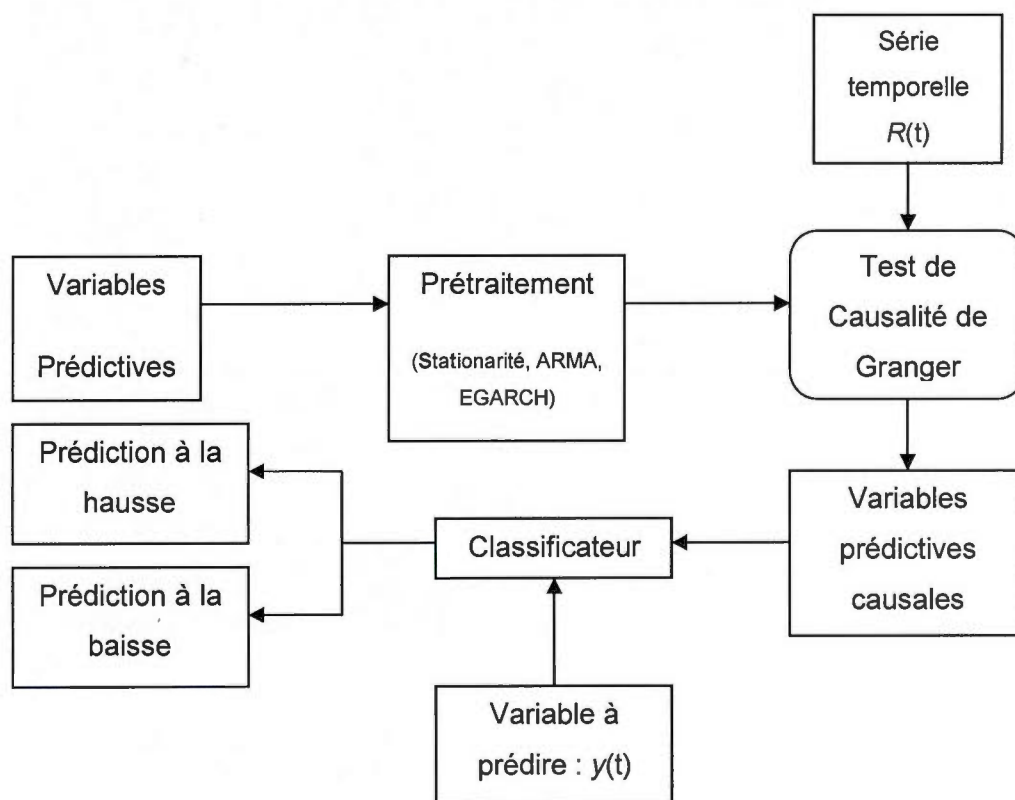


Figure 3.14 Schéma général du système prédictif

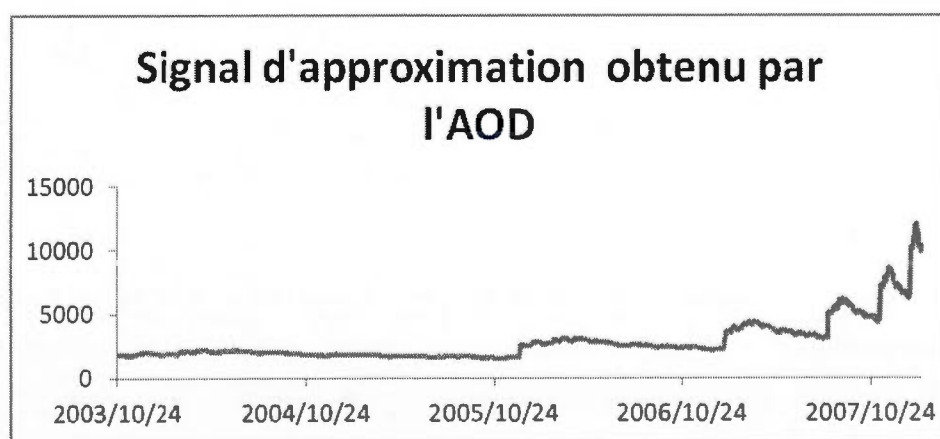


Figure 3.15 Signal de basses fréquences obtenu par AOD

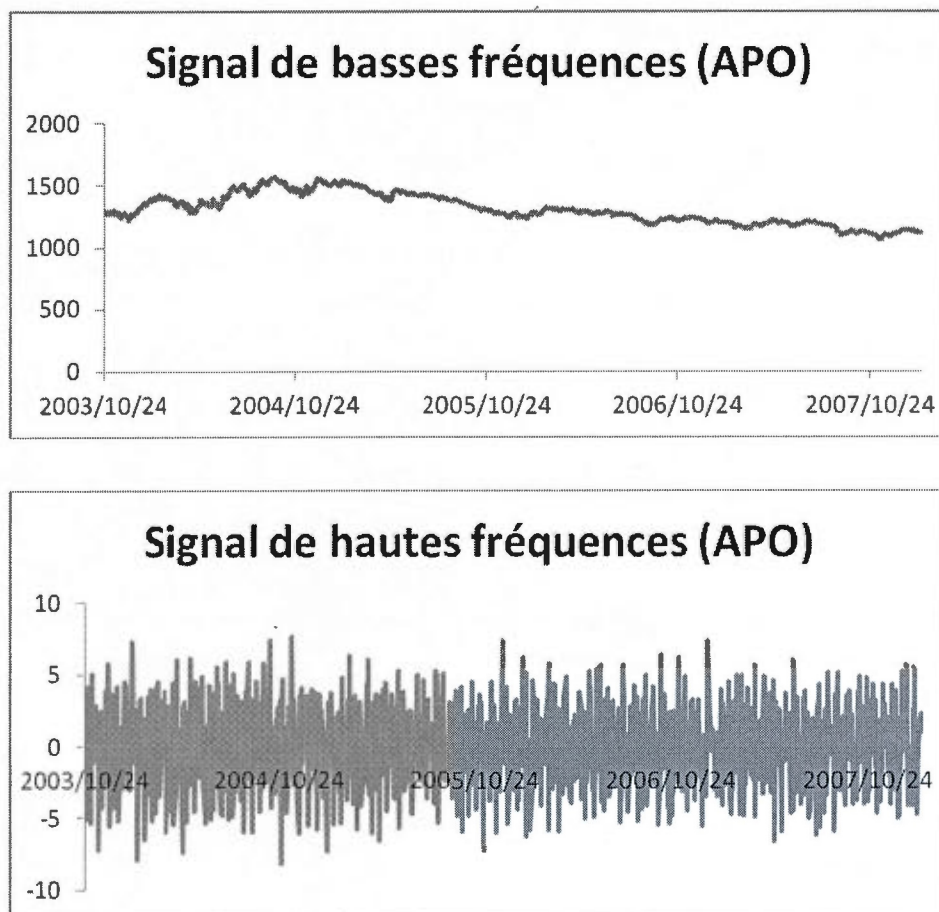


Figure 3.16 Décomposition obtenue par APO

Dans le chapitre suivant, les résultats des simulations pour chaque catégorie d'information avant validation croisée sont présentés. De plus, nous présenterons les résultats du meilleur modèle (couple information-classifieur) avant et après validation croisée. Finalement, l'effet du nombre de couches cachées sur la performance du meilleur modèle sera examiné.

CHAPITRE IV

LES RÉSULTATS

Ce chapitre est composé de trois sections. Dans la première, nous présentons les résultats obtenus pour chaque catégorie d'information avant validation croisée. Dans la deuxième section, les résultats des meilleurs modèles (couples information-classifieur) sont présentés avec validation croisée. De plus, nous examinons l'effet du nombre de couches cachées sur la performance du meilleur modèle retenu à la fin de l'étude. Finalement, une conclusion partielle est donnée au niveau de la dernière section.

4.1 Résultats sans validation croisée

Avant de présenter les résultats, nous citons les variables d'entrées identifiées par le test de causalité de Granger (1969). Au niveau des variables économiques, le test de causalité sélectionne DFF (Effective Federal Funds Rate), DEXSZUS (taux de change Switzerland / U.S.), DTWEXB (Trade Weighted Exchange Index: Broad), et DXJPUS (taux de change Japan/U.S.). Au niveau des variables techniques, le test de causalité sélectionne les indicateurs X4 et EMV. Finalement, au niveau des mesures de sentiment du marché boursier, le test de Granger identifie la volatilité des changements ($R(t)$) de l'indice boursier S&P500, et l'État psychologique du marché représenté par l'indice ICI (Investor Confidence Index). Les résultats sont présentés dans le Tableau 4.1 où la statistique probabilité est associée à l'acceptation ou au rejet de l'hypothèse nulle.

Tableau 4.1

Variables prédictives sélectionnées par le test de causalité

Hypothèse nulle	Probabilité
Variables Économiques	
SDFF ne cause pas R	0.00645
SDEXSZUS ne cause pas R	0.02884
SDTWEXB ne cause pas R	0.04953
SDXJPUS ne cause pas R	0.03735
Variables Techniques	
X4	0.0038
EMV	0.00918
Mesures de sentiment	
Volatilité de R ne cause pas	
R	0.01628
ICI ne cause pas R	0.00578

L'introduction au test de causalité de Granger (1969) et les détails des tests sont donnés en annexe D. Il faudrait noter⁴¹ que les entrées SDFF, SDEXSZUS, SDTWEXB, SDXJPUS sont respectivement les différences premières de DFF, DEXSZUS, DTWEXB, DXJPUS. Finalement, la modélisation ARMA(p,q) de la série des variations ($R(t)$) de l'indice boursier montre que le processus suivi par $R(t)$ est un ARMA(2,0) ou encore AR(2). Par conséquent, le nombre de retards est deux ($R(t), R(t-1)$) pour former les entrées au RNA et au SVM selon l'approche qui utilise de l'information dite historique pour prédire $y(t+1)$ la tendance future du marché. Ce résultat a aussi été confirmé par l'examen de la fonction d'auto corrélation partielle. Les détails de ces résultats sont donnés en annexes E.

⁴¹ En effet, elles ont toutes subies une différence première car elles n'étaient pas stationnaires selon le test ADF. Par conséquent, le test de causalité de Granger (1969) ne pourrait être appliqué qu'après avoir effectué cette transformation linéaire (voir les annexes C et F). En d'autres mots, le test de causalité de Granger (1969) suppose que les séries en questions soient stationnaires en moyenne.

Rappelons que le nombre de neurones dans la couche cachée est fixé au même nombre (Exact), au double et au triple de neurones dans la couche d'entrée. Le nombre des couches cachées est fixé à un puisque le RNA standard est un approximateur universel de fonctions (Cybenko, 1989; Funahashi, 1989; Hornik, 1991). De plus, seuls les meilleurs résultats obtenus pour chaque type d'information seront présentés⁴² pour mieux synthétiser l'information et présenter les résultats pertinents.

Avant de présenter les résultats de nos simulations, les questions de la thèse auxquelles nous voudrions répondre sont citées en rappel :

Question.1 :

Quel type d'information (économique, technique, historique, sentiment) pourrait fournir les meilleures prédictions ?

Question.2 :

Est ce qu'une combinaison des différents types d'information permettrait d'améliorer les prévisions ?

Question.3 :

Quel algorithme d'approximation numérique permet un meilleur apprentissage au RNA et donc de mieux modéliser la série temporelle et fournir les meilleures prédictions ?

⁴² En effet, il s'agit de réaliser un grand nombre d'expériences dont les résultats doivent être synthétisés en ne présentant que les meilleurs modèles. Par exemple, pour le RNA 105 simulations vont être faites : 5 algorithmes \times 7 types d'informations \times 3 modes de variations du nombre de neurones dans la couche cachée (le nombre de neurones que dans la couche d'entrée, ou le double ou le triple). Pour le SVM, le nombre de simulations est 21 : 7 catégories d'informations \times 3 ordres de la polynomiale. Donc au total, 126 expériences sont réalisées dans le cadre de cette thèse.

Question 4 :

Est ce qu'on peut améliorer les résultats par rapport aux catégories d'information traditionnelles en considérant de l'information temps-fréquence en utilisant l'analyse multi-résolution; en l'occurrence l'analyse par paquets d'ondelettes ?

Les résultats des simulations présentés dans la Figure 4.1 montrent que l'information de type économique performe mieux avec les algorithmes numériques Powell-Beale (Triple : le triple des neurones dans la couche cachée) en obtenant 57.94% de taux de classification (détection de tendance future) correcte. Par ailleurs, les algorithmes numériques LM (Exact) et BFGS (Doubles) ont obtenu le même taux de classification (56.54%). L'algorithme de Fletcher-Reeves (Exact) a performé le moindre par rapport à tous les autres algorithmes numériques : 55.14% de taux de prédiction correcte. Malgré le fait que les taux de classification sont clairement bas, les algorithmes d'entraînement Powell-Beale, LM, BFGS, Polak-Ribière, et Fletcher-Reeves ont obtenu des taux de sensibilités très encourageants allant de 85% à 93.33%. Donc, la détérioration des taux de classification vient en fait des taux de spécificité qui sont très bas. Finalement, le SVM a réalisé la plus basse performance en termes de taux de classification correcte (53.27%) et en termes de sensibilité (77.5%).

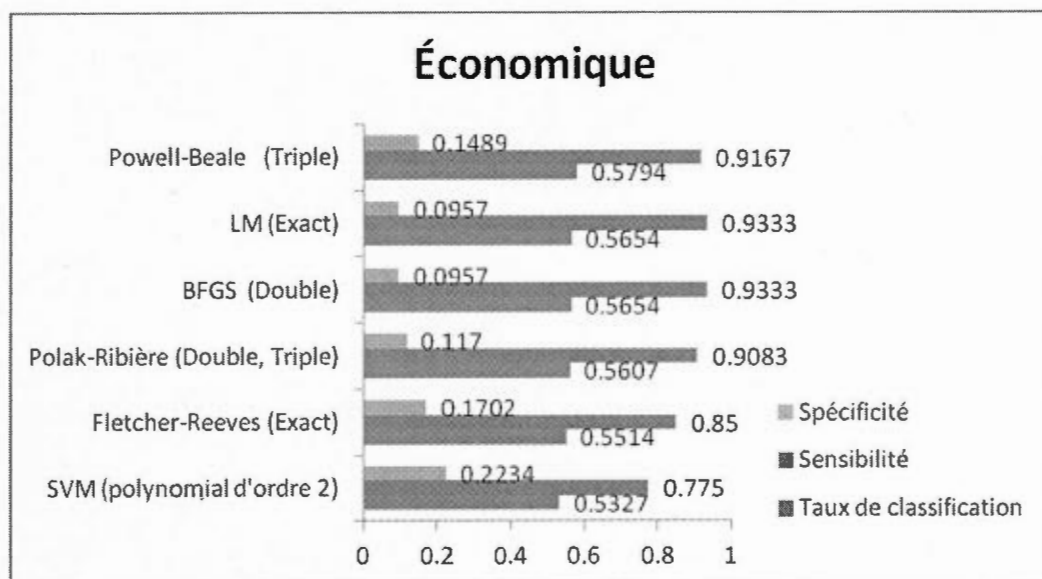


Figure 4.1 Résultats des simulations : variables économiques

Les résultats des simulations présentés dans la Figure 4.2 montrent que l'information de type technique performe mieux avec les algorithmes numériques Polak-Ribière (Exact), Fletcher-Reeves (Double), et Powell-Beale (Triple). Ils ont tous atteint 57.01% de taux de classification correcte. Ils sont suivis par le SVM à polynomial d'ordre deux qui a atteint 55.14% de taux de détection correcte, et du LM (Exact) et du BFGS (Double) qui ont obtenu tous les deux un taux de détection correcte de l'ordre 54.21%. Le taux de sensibilité demeure haut pour tous les classifieurs (87.5% à 95%). Par ailleurs, le taux de spécificité est remarquablement très bas : 5.32% à 11.7%.

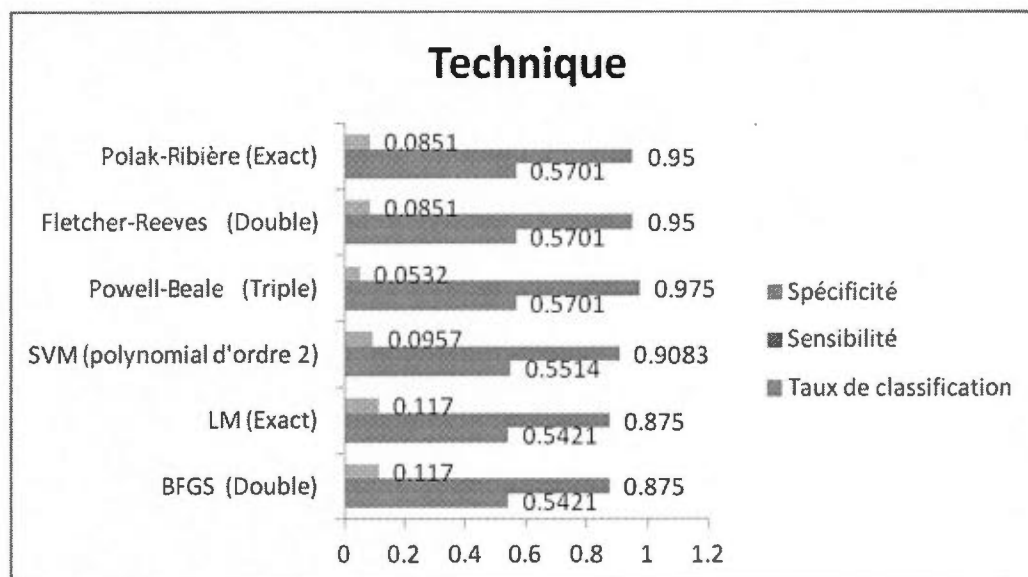


Figure 4.2 Résultats des simulations : variables techniques

Les résultats des simulations présentés dans la Figure 4.3 montrent que l'information de type historique performe mieux avec les algorithmes numériques Polak-Ribière (Triple) et LM (Double) qui ont atteint respectivement 58.41% et 57.01% de taux de classification correcte. Les algorithmes numériques de BFGS (Double et Triple), Fletcher-Reeves (Exact), et Powell-Beale (Exact et Double) ont tous obtenu 56.07% de taux de détection correcte. Ces derniers algorithmes d'entraînement ont atteint 100% de sensibilité et 0% de spécificité. Finalement, le SVM à polynomial d'ordre deux a réalisé un taux de classification correcte de 56.34%. Donc sa performance est bien meilleure que celles obtenues par le BFGS, Fletcher-Reeves, et Powell-Beale. De plus, il a réalisé la meilleure spécificité (61.7%) comparativement à tous les RNA. Finalement, les taux de spécificité obtenus en utilisant de l'information de type historique sont bas (0% à 61.17%); mais les taux de sensibilité sont clairement plus élevés : 50.83% à 100%. Encore une fois, les résultats montrent que les classifieurs ont réussi à obtenir des taux de sensibilité supérieurs.

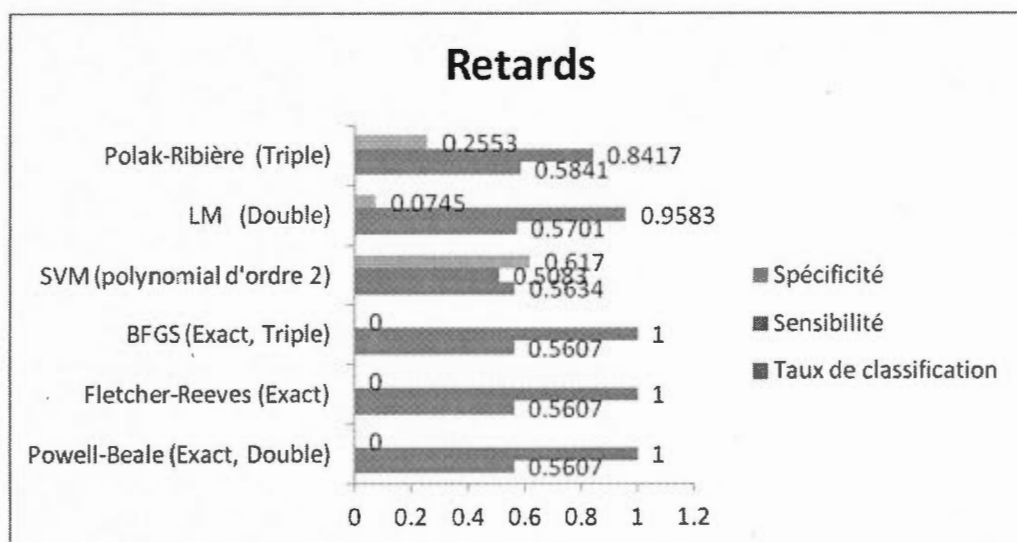


Figure 4.3 Résultats des simulations : approches historiques de la série

Les résultats des simulations présentés dans la Figure 4.4 montrent que l'information de type sentiment (État psychologique du marché) performe mieux avec le SVM à polynomial d'ordre 3 et le BFGS (Triple) qui ont obtenu un taux de détection correcte de l'ordre de 58.41%. Les algorithmes LM (Double) et Polak-Ribière (Double) ont réalisé les taux de classification les plus bas (52.8% et 51.87%; respectivement) mais ils ont obtenu les taux de spécificité les plus hauts : 32.98% et 47.87% respectivement. Finalement, les résultats montrent que le SVM à polynomial d'ordre 3, le BFGS, LM, et Polak-Ribière ont réussi à atteindre des taux de sensibilité élevés : 90.83% à 100%.

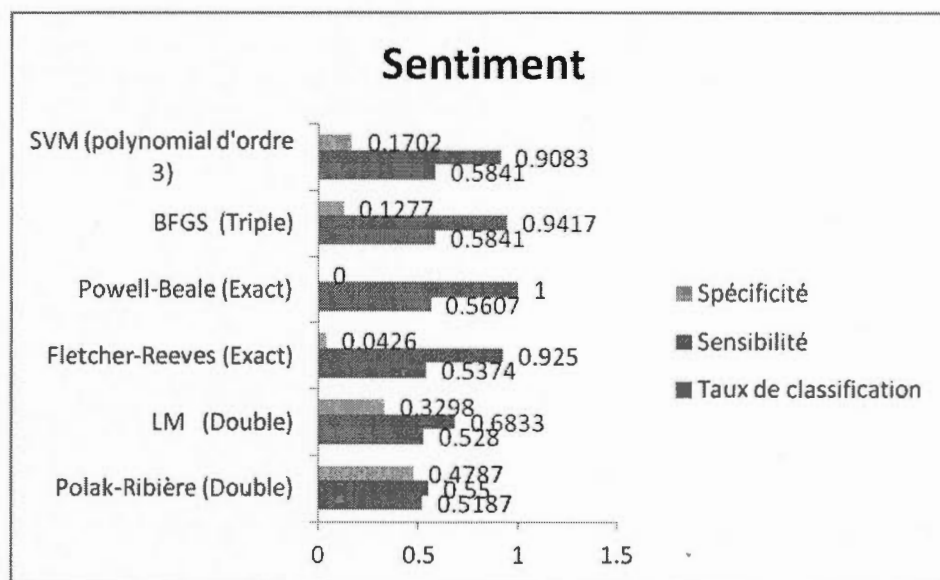


Figure 4.4 Résultats des simulations : variables de sensiment

Les résultats des simulations présentés dans la Figure 4.5 montrent que l'information d'ensemble performe mieux avec l'algorithme numérique Fletcher-Reeves (Triple). Par ailleurs, elle performe moins en utilisant l'algorithme de Powell-Beale (Double) et le SVM à polynomial d'ordre deux. Ils ont atteint respectivement 50.47% et 51.68% de taux de détection correcte. Trois faits marquants sont montrés la Figure 30 par rapport aux résultats montrés dans les figures précédentes. Les taux de sensibilité ont diminué et les taux de spécificité se sont améliorés. Par ailleurs, les taux de classification correcte sont en général similaires à ceux obtenus avec les autres types d'information.

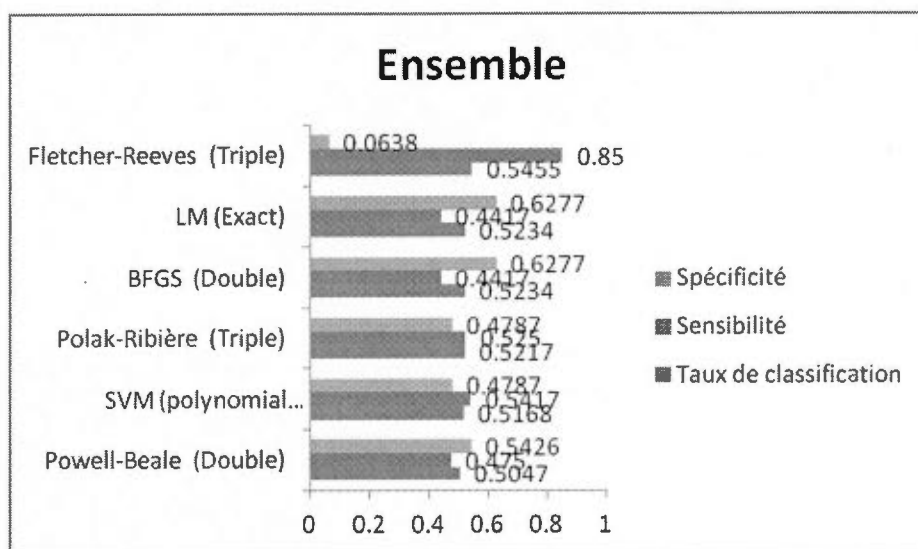


Figure 4.5 Résultats des simulations : ensemble des informations

Les résultats des simulations présentés dans la Figure 4.6 montrent que tous les algorithmes numériques étudiés dans le cadre de cette thèse pour l'entraînement des RNA performant de manière égale en utilisant de l'information extraite par analyse par ondelette discrète (AOD) suivant la méthodologie de Wang (2011). Le taux de classification correcte, la sensibilité, et la spécificité obtenue sont respectivement 56.07%, 100% et, 0%. L'approche par AOD montre que tous les algorithmes ont réussi à prédire à 100% toutes les tendances haussières. Nous rappelons le lecteur que dans notre définition du problème, la sensibilité mesure le taux d'erreur de classification des tendances en baisse; par exemple prévoir une vraie baisse du marché comme étant une hausse éventuelle. La spécificité mesure le taux d'erreur de classification des tendances en hausse; par exemple prévoir une vraie hausse du marché comme étant une baisse éventuelle. De ce fait, les algorithmes étudiés ont tous échoué à prédire les tendances baissières du marché boursier. Finalement, les résultats montrent que le SVM à polynomial d'ordre 2 a atteint 49.25% (taux de détection correcte), 44.17% (sensibilité), et 53.19% (spécificité).

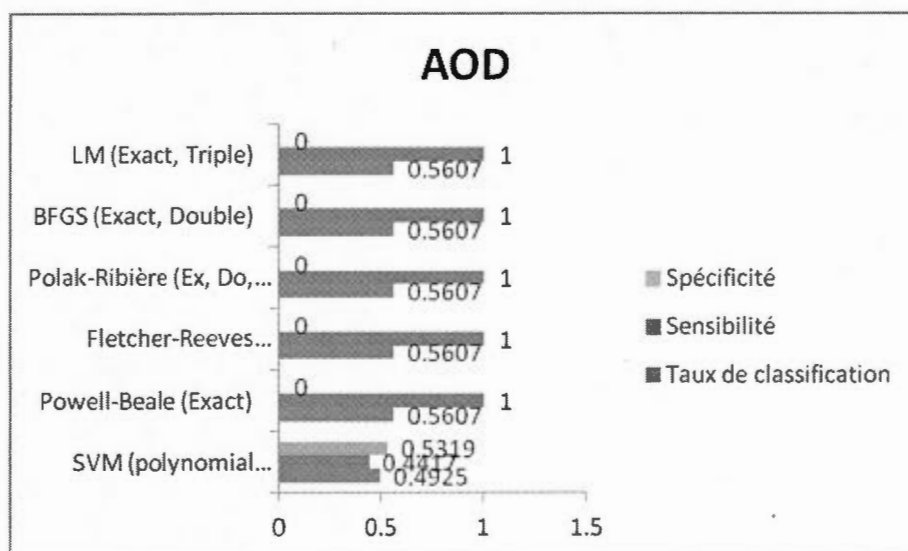


Figure 4.6 Résultats des simulations : analyse par ondelette discrète

Finalement, les résultats des simulations présentés dans la Figure 4.7 montrent que l'information de type APO permet clairement d'améliorer le taux de prédiction de tous les classifieurs par rapport à toutes les autres informations classiques (économique, technique, historique, psychologique). La meilleure performance a été obtenue par les algorithmes numériques LM (Double) et Powell-Beale (Exact); soit 72.43%. Par contre, les moins performant sont les algorithmes numériques BFGS (Exact, Triple) et Fletcher-Reeves (Exact) qui ont tout les deux atteint 71.96% de taux de détection correcte.

Finalement, le SVM à polynomial d'ordre 3 a atteint 71.96% de taux de classification correcte. De plus, la Figure 4.7 montre que la spécificité s'est largement améliorée pour tout les classifieurs; allant de 53.19% à 75.53%. Ce résultat est largement supérieur à ce qui a été observé dans les Figures 4.1 à 4.6.

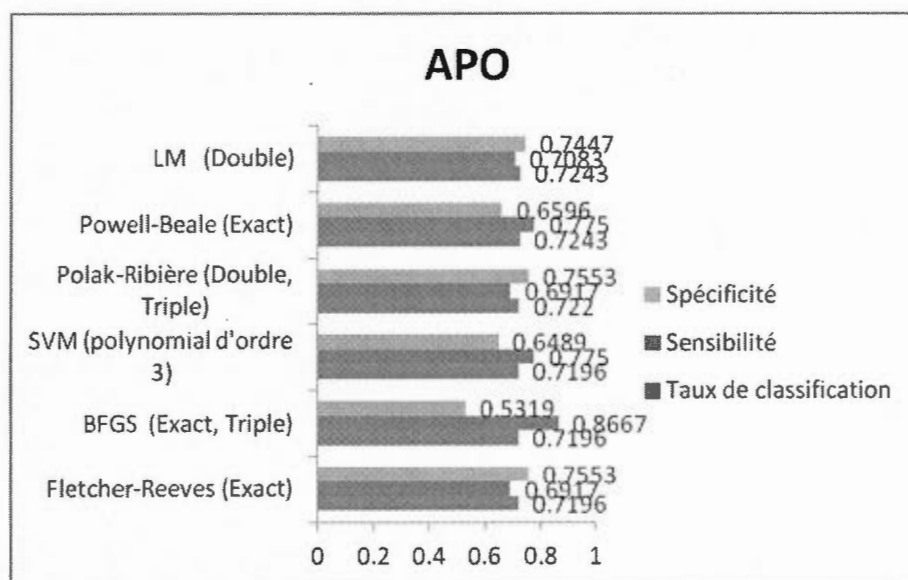


Figure 4.7 Résultats des simulations : analyse par paquet d'ondelettes

Pour mieux synthétiser l'ensemble des résultats montrés dans les Figures 4.1 à 4.7, nous avons produit la Figure 4.8 qui résume l'ensemble de l'information (hors analyse par ondelette) pour pouvoir répondre aux questions 1, 2 & 3 mentionnées au début de ce chapitre. En effet, la Figure 4.8 présente la meilleure performance obtenue pour chaque catégorie d'information utilisée pour prédire la tendance à la hausse ou à la baisse du marché boursier S&P. En d'autres mots, cette figure fait un classement des meilleures combinaisons classifieur-information selon le taux de classification correcte par ordre décroissant.

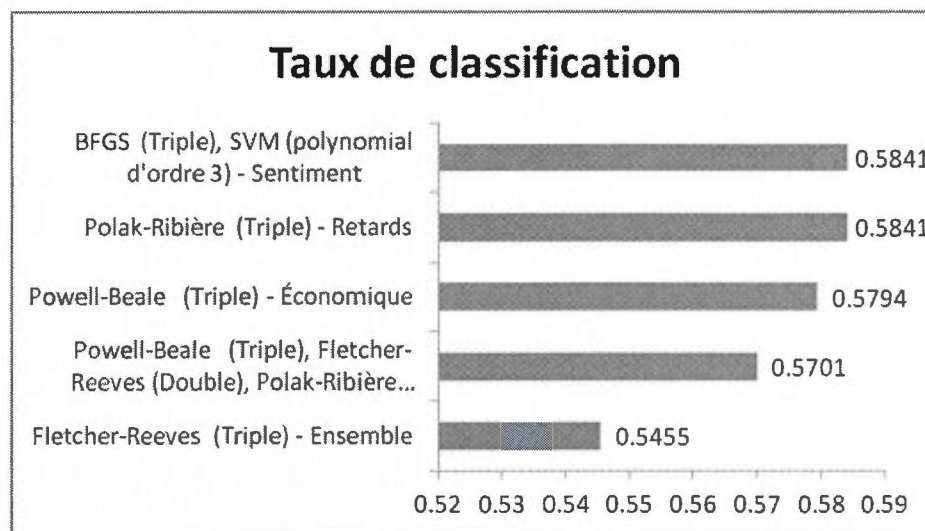


Figure 4.8 Meilleurs performances selon le type d'information/classifieur

Réponse à la Question.1 :

La Figure 4.8 montre que l'information sur l'État psychologique du marché boursier mesurée par les variables de sentiment et la catégorie d'information de type historique représentent les meilleurs types d'informations classiques pour prédire le marché S&P500 sur la période étudiée : 2003-2007. Ces deux catégories ont obtenu un taux de classification égale à 58.41%.

Réponse à la Question.2 :

Les résultats (Figure 4.8) montrent que l'algorithme de Fletcher-Reeves (Triple) qui a obtenu le meilleur taux de classification (54.55%) parmi les autres algorithmes numériques d'entraînement du RNA en utilisant de l'information d'ensemble; se classe dernier en termes de taux de classification par rapport aux autres classifieurs qui ont utilisé différentes catégories d'information classique. De ce fait, la combinaison des différentes catégories d'information n'améliore pas le taux de classification.

Réponse à la Question.3 :

Les Figures 4.1 à 4.7 montrent en général qu'il n'y a pas de différences majeures entre les performances des algorithmes numériques étudiés dans le cadre de cette thèse dans la prédiction de la tendance future du marché boursier S&P500. Cependant, à la lumière de la Figure 4.8 on pourrait dire que les algorithmes de BFGS et Polak-Ribière performant le mieux lorsqu'ils utilisent respectivement de l'information de type sentiment du marché et données historiques (retards). Mais en général, quoique les performances des algorithmes numériques étudiés soient similaires en termes de taux de classification correcte, on peut dire que Polak-Ribière performe "généralement" bien (Figures 4.1-4.7).

Réponse à la Question.4 :

Selon la Figure 4.9 de synthèse, l'information extraite par APO permet d'améliorer remarquablement le taux de classification; et donc de prédiction des tendances haussières et baissière du marché boursier S&P sur la période 2003-2007. Par exemple, la différence entre la performance obtenue par APO (72.43%) et les mesures de l'État psychologique du marché (58.41%) est de l'ordre de 14.08 points de bases. Par ailleurs, la différence entre la performance obtenue par l'APO et l'ensemble des catégories d'informations (54.55%) est de l'ordre de 17.88 points de bases.

Donc, on peut effectivement améliorer les résultats par rapport aux catégories d'information traditionnelles en considérant de l'information temps-fréquence obtenue par l'analyse multi-résolution en utilisant l'analyse par paquets d'ondelettes.

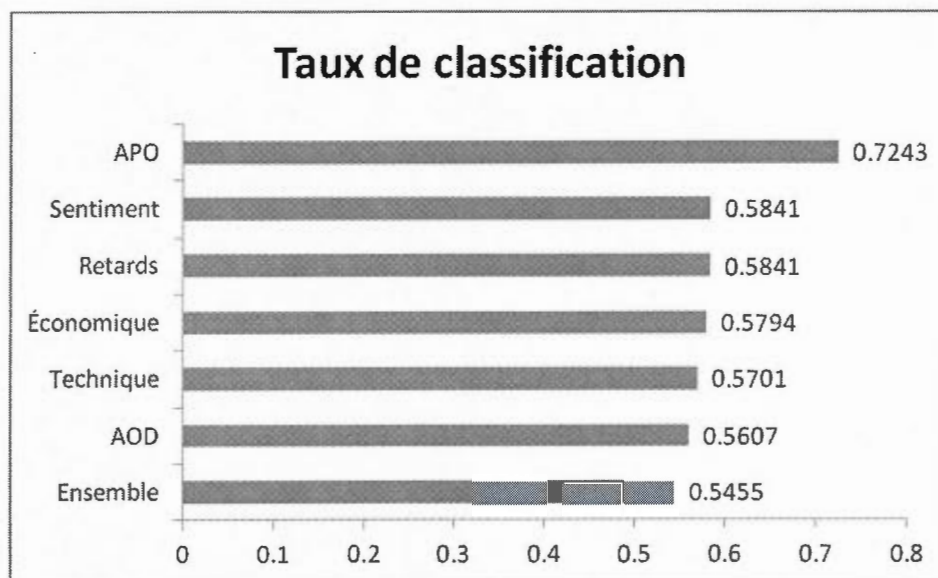


Figure 4.9 Synthèse des résultats par type d'information

4.2 Résultats avec validation croisée

Avant de conclure ce chapitre, nous voulons examiner l'impact du nombre de couches cachées dans le RNA entraîné par LM en utilisant de l'information obtenue par APO sur le taux de classification correcte; et nous voulons également présenter les résultats obtenus par la technique de validation croisée à 10 blocs aléatoires. Comme nous l'avons déjà mentionné au niveau de la section 2.6, Il n'est pas pertinent d'effectuer une validation croisée et donc calculer la moyenne et l'écart type du taux de classification parce que nous faisons de la prédiction d'une série temporelle et donc nous ne pouvons pas utiliser le futur pour prédire le passé. En effet, il est important de préserver la chronologie des événements car elle possède de l'information sur la dynamique des entrées et de la variable à prédire. Par contre la validation croisée sera certainement utile dans le cadre de la classification d'images ou d'objets pour lesquels le facteur temps n'existe pas. En somme, il s'agit d'un problème de classification dans le contexte d'analyse de séries chronologiques dont le facteur temps est déterminant de la dynamique des informations qui circulent.

En effet, les études qui traitent de la classification des séries temporelles n'ont pas adopté la validation croisée (Fernandez-Rodriguez et al., 2000; Harvey et al., 2000; Tay et Cao, 2001; Halliday, 2004; Doesken et al., 2005; Huang et al., 2005; Atsalakis et Valavanis, 2009; Azzini et al., 2010; Cengiz et Herwartz, 2011; Nyberg, 2011; Wang et al., 2011). Par conséquent, nous allons commenter brièvement les résultats obtenus par validation croisée et nous laissons le soin au lecteur d'en tirer les conclusions. En somme, l'objectif est de rapporter ces résultats à titre indicatif et le lecteur sera en mesure d'évaluer leur pertinence. La Figure 4.10 montre les résultats des meilleurs modèles présentés dans les Figures 4.8 et 4.9 après avoir effectué une validation croisée à dix blocs aléatoires. Pour chaque modèle (information-classifieur) la moyenne et l'écart type du taux de classification correct sont calculés. En somme, la Figure 4.10 montre deux faits marquants. Le premier est que les résultats obtenus sont nettement similaires aux résultats obtenus sans validation croisée (Figure 4.8 et 4.9). Le deuxième point est que l'approche par analyse en paquet d'ondelette continue toujours à être supérieure par rapport aux autres catégories d'information; soit $72.85\% \pm 0.0324$ en utilisant l'algorithme de Powell-Beale pour entraîner le RNA et $72.52\% \pm 0.0241$ en utilisant l'algorithme LM.

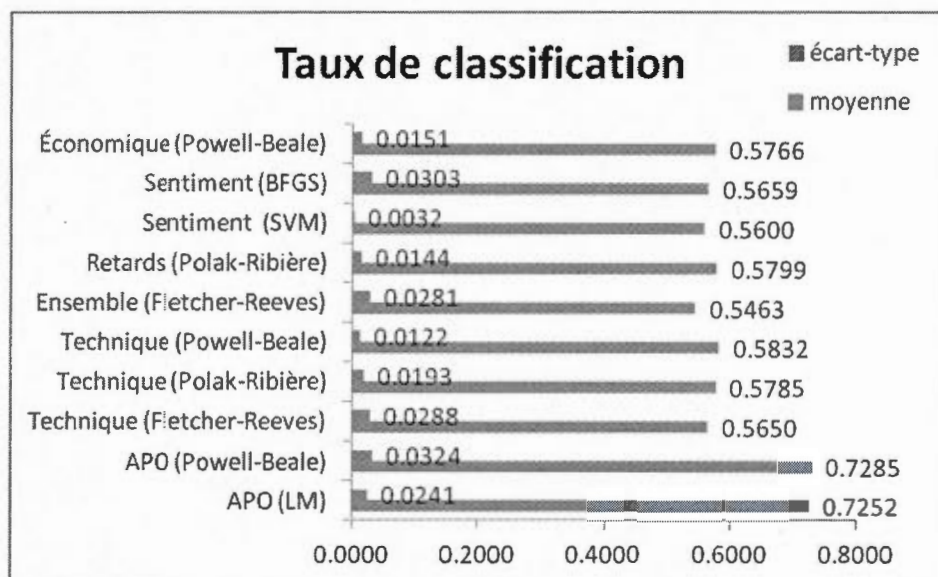


Figure 4.10 Résultats de la validation croisée

Comme nous l'avons dit au niveau de la section 2.3, pour le meilleur modèle obtenu (le couple algorithme numérique et information qui obtient la meilleure performance de classification) nous allons faire varier le nombre de couches cachées pour vérifier s'il y a possibilité d'améliorer davantage sa performance déjà obtenue avec une seule couche cachée. Donc, selon les résultats obtenus (Figure 4.10) le RNA entraîné par l'algorithme de LM (Double)⁴³ et le RNA entraîné par l'algorithme de Powell-Beale (Exact)⁴⁴ qui utilisent de l'information extraite par APO sont les meilleurs modèles pour prédire la tendance future du marché. Pour examiner l'impact de nombre de couches sur la performance de ces deux modèles nous allons varier le nombre de couches cachées de 1 à 4. Dans chaque cas, les nombre de neurones dans la cache

⁴³ Double : nombre de neurones dans la couche caché est le double de neurones dans la couche d'entrée.

⁴⁴ Exact : nombre de neurones dans la couche caché est le même que le nombre de neurones dans la couche d'entrée.

cachée reste fixe : double et exact pour les topologies des RNA entraînés par LM et Powell-Beale respectivement. Les résultats sont présentés dans la Figure 4.11.

Elle montre clairement que le fait de mettre plus d'une couche cachée dans le réseau ne permet pas d'améliorer le taux de classification (détection) des tendances haussières et baissières du marché S&P. Ce résultat relatif à l'effet du nombre de couches cachées sur la performance du réseau de neurones est consistant avec la littérature tant au niveau théorique (Cybenko, 1989; Funahashi, 1989; Hornik, 1991; Fausett, 1994) qu'au niveau empirique (Burkitt, 1991; Kavzoglu, 1999; Shafi et al., 2006).

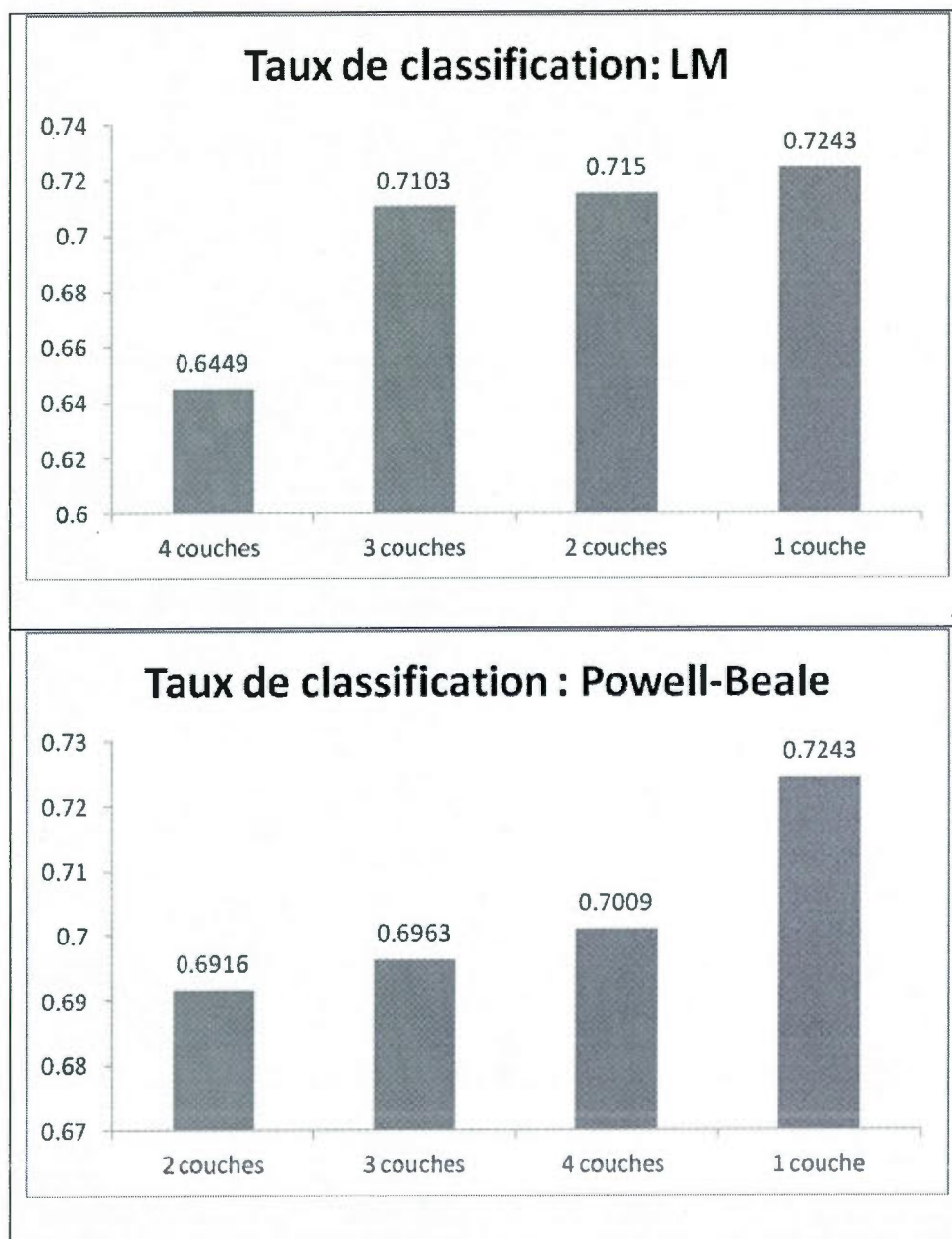


Figure 4.11 Effet du nombre de couches cachées sur la performance

4.3 Conclusion partielle

En termes de conclusion de ce chapitre, nous avons répondu aux questions 1, 2,3, et 4 comme suit :

- 1) Les informations de types historique et psychologique sont plus pertinentes pour la prédiction de la tendance (haussière et baissière) future du S&P500.
- 2) La combinaison des différentes catégories d'information (économique, technique, historique, sentiment) n'améliore pas la précision de la prédiction de la tendance future du marché S&P500.
- 3) Le meilleur algorithme d'entraînement du RNA pour le S&P500 est Polak-Ribière ayant pour entrée de l'information historique et BFGS ayant pour variables prédictives les mesures de sentiment.
- 4) L'information fréquentielle extraite par l'analyse multi-résolution par paquet d'ondelettes (APO) permet largement d'améliorer la performance prédictive des RNA en termes de taux de classification correcte quelque soit l'algorithme numérique utilisé pour l'entraînement. À ce niveau, tous les classifieurs ont bien performés; en particulier les RNA entraînés avec les algorithmes numériques LM et Powell-Beale.

De plus, quelques points importants doivent être mentionnés. Tout d'abord, bien que les résultats obtenus par les RNA varient selon le type d'information et l'algorithme d'apprentissage il faudrait noter qu'en général que les différences en termes de performances (taux de classification) sont minimales. Par exemple, pour une catégorie d'information donnée, les taux de classification obtenus par les différents algorithmes d'entraînement sont très similaires. De plus, les performances des RNA dépendent du nombre de neurones dans la couche cachée. En outre, l'augmentation du nombre de couches cachées n'améliore pas le taux de classification. En d'autres mots, une seule

couche cachée dans le réseau est suffisante pour mieux prédire la tendance future du marché S&P.

Le deuxième point intéressant est que la sensibilité des classifieurs (RNA et SVM) est généralement élevée et ce pour toutes les informations conventionnelles (économique, technique, historique, psychologique) utilisées par ces machines. Par contre, le taux de spécificité est généralement trop bas et même nuls dans certains cas.

Par conséquent, quelque soit l'information conventionnelle utilisée par le classifieur; ce dernier ne détecte pas correctement les tendances baissières dans la plus part des cas. Cependant, la spécificité des classifieurs s'améliore clairement avec l'usage de l'information obtenue par l'analyse en paquet d'ondelette. Finalement, les simulations ont montré qu'en général les RNA ont performé mieux que le SVM quoique les différences de performances entre ces deux classifieurs soient minimales.

Le troisième point est que les résultats obtenus par validation croisée confirment les résultats obtenus sans validation croisée. Cependant, il faudrait être prudent avec les résultats de la validation croisée parce que les données en question sont des séries temporelles et en faisant de la validation croisée la dynamique de l'information contenue dans les données a été interrompue. Donc, les résultats obtenus par validation croisée sont à considérer avec précaution.

Maintenant que nous avons présenté les résultats avant et après simulations par validation croisée et que nous avons répondu aux questions de la thèse avec certaines observations, nous allons discuter les résultats obtenus au niveau du chapitre suivant.

CHAPITRE V

DISCUSSION DES RÉSULTATS

Dans ce chapitre nous allons discuter chaque question à laquelle nous avons répondu. De plus, nous discuterons de la raison pour laquelle le RNA performe mieux que le SVM. Finalement, nous comparerons nos résultats obtenus à ceux de la littérature. À ce niveau, nous discuterons du niveau du taux des reconnaissances obtenus des éventuelles hausses et baisses du marché S&P500. Tout d'abord, nous voudrions faire un rappel des questions de la thèse.

Question.1 :

Quel type d'information (économique, technique, historique, sentiment) pourrait fournir les meilleures prédictions ?

Question.2 :

Est ce qu'une combinaison des différents types d'information permettrait d'améliorer les prévisions ?

Question.3 :

Quel algorithme d'approximation numérique permet un meilleur apprentissage au RNA et donc de mieux modéliser la série temporelle et fournir les meilleures prédictions ?

Question 4 :

Est ce qu'on peut améliorer les résultats par rapport aux catégories d'information traditionnelles en considérant de l'information temps-fréquence en utilisant l'analyse multi-résolution; en l'occurrence l'analyse par paquets d'ondelettes ?

Première question

Les simulations ont clairement montré que les données sur l'état psychologique du marché et sur l'historique des rendements permettent de mieux prédire la tendance (hausse ou baisse) du S&P500. En d'autres mots, les mesures de sentiments des investisseurs et les retards (mémoire courte de la série temporelle) sont plus pertinents que les informations techniques et économiques. Ce résultat est expliqué à plusieurs niveaux. D'abord, la mesure de sentiment représentée par l'indice VIX reflète bien le risque "futur" du marché estimé par la volatilité des options sur l'indice S&P500 (Whaley, 2000; Bandopadhyaya et Jones 2008; Ben-Rephael et al., 2010). Par ailleurs, la volatilité du marché estimée par le modèle EGARCH représente le risque contemporain du marché et elle est très associée aux rendements des titres (Barber et al.; 2009, Yang et Wu, 2011; Yu et Yuan, 2011 ; Lawrence et Makridakis, 1989 ; Yu et Yuan, 2011 ; Wang et al., 2006; Schmeling, 2007; Verma et Verma, 2007). D'ailleurs, les résultats des tests de Granger (1969) confirment les conclusions de ces études précédentes. Ensuite, l'historique de la série temporelle de l'indice boursier permet de détecter une sorte de mémoire à court terme qui caractérise la série temporelle en question. De ce fait, l'information basée sur l'historique des mouvements du marché permet de mieux prédire sa tendance future que l'information économique et technique. Par exemple, nous avons trouvé que le rendement actuel ($R(t)$) est auto-corrélé avec les rendements antécédents à deux retards : $R(t-1)$ et $R(t-2)$. C'est-à-dire que le prix d'aujourd'hui est influencé par l'information contenue dans le prix d'hier et avant-hier. De ce fait, l'historique du comportement de l'indice boursier S&P500 a réussi à prédire la tendance future de ce

dernier mieux que l'information technique ou l'information économique. Ce résultat confirme l'importance de la mémoire courte dans les séries temporelle boursières dans la modélisation et la prédiction (Lo et MacKinlay, 1988; Poterba et al., 1988; Cutler et al., 1990; Mills, 1993; Cheung et Lai, 1995; Grang et et Hyung, 2004; Mukherjee et al., 2011; Rege et Martin, 2011).

Deuxième question

Nous avons trouvé que lorsque l'information économique, technique, historique, et sentiment composent un ensemble d'entrées au RNA, la performance de ce dernier se dégrade considérablement. Ce résultat a été également confirmé par le SVM. La contreperformance du RNA en utilisant l'ensemble de toutes les informations disponibles est expliquée par le fait que les différentes catégories d'information qui composent cet ensemble ne sont pas similaires. Une manière de réduire l'effet de la dissimilarité serait d'employer un réseau de type carte auto-organisée qui permet d'effectuer un groupement (clustering) des entrées selon leur degré de similarité. Cependant, cette approche ne permet pas d'identifier quelles entrées faudrait-il retenir. Une manière plus intelligente d'approcher ce problème serait de concevoir un système d'ensemble (comité machine) composé de plusieurs RNA; dont chaque RNA est spécialisé dans l'apprentissage d'une seule sorte d'information. Ensuite, la décision finale serait agrégée par un autre RNA. L'avantage de ce type de système prédictif est que la décision finale est prise par le RNA de sortie qui choisi intelligemment l'information la plus appropriée pour prédire une éventuelle hausse ou baisse. En effet, cette approche fera l'objet des travaux futurs.

Troisième question

Les résultats obtenus montrent qu'en général les algorithmes basés sur le gradient conjugué (en particulier Polak-Ribière) réalisent de meilleures performances (Figures 4.1-4.9) que l'algorithme de BFGS (quasi-Newton) et l'algorithme de Levenberg–

Marquardt (L-M) qui – ce dernier- est le plus utilisé en finance à cause de sa simplicité et sa rapidité de convergence. Les algorithmes L-M et BFGS utilisent l'information contenue dans la Hessienne pour effectuer les déplacements appropriés. Par contre, les algorithmes du gradient conjugué utilisent implicitement l'information contenue dans la Hessienne tout en effectuant une recherche de la nouvelle direction. En particulier, la méthode du gradient conjugué cherche d'abord à trouver une direction conjuguée et ensuite performe une optimisation unidimensionnelle selon cet axe.

De plus, la principale caractéristique des algorithmes de type gradient conjugué est l'introduction d'un effet de mémoire pour le calcul efficace de la future direction de recherche : les algorithmes du gradient conjugué sauvegardent en mémoire l'ancien gradient et l'ancienne direction de recherche afin de trouver la nouvelle direction.

Nous avons trouvé que l'information extraite par l'analyse en paquet d'ondelette est la meilleure catégorie pour la prédiction de la tendance future du marché. Les résultats montrent également qu'en général (Figures 4.3, 4.4, 4.6, 4.7) l'algorithme de Polak-Ribière est le plus approprié pour l'entraînement du RNA. Ceci est expliqué par l'introduction de l'effet de mémoire – mentionné plus haut – et surtout que l'optimisation devient plus robuste en général sous Polak-Ribière parce que le choix de la nouvelle direction de recherche est modifié de façon adéquate par rapport à Fletcher-Reeves. Par conséquent, Polak-Ribière est plus efficace que Fletcher-Reeves pour l'optimisation de fonction non quadratique (Bertsekas, 1999). En d'autres mots, Polak-Ribière est plus général que Fletcher-Reeves. En effet, la forme de la relation entre les entrées et les sorties est supposée être non linéaire mais pas nécessairement quadratique. De ce fait, Polak-Ribière permet de donner de meilleurs résultats.

Quatrième question

Au niveau de l'analyse fréquentielle par ondelette, les résultats des simulations ont montré que l'analyse par la transformée en ondelette discrète (AOD) n'améliore pas le taux de prédiction du RNA et du SVM par rapport à l'information économique. Cependant, l'usage de l'information extraite par l'analyse par paquet d'ondelette (APO) permet aux RNA et au SVM d'atteindre leurs performances maximales; à savoir 72.43% et 71.96% respectivement contre 58.17% et 56.34% en employant de l'information historique et 58.41% pour les deux en utilisant de l'information sur l'état psychologique du marché. Ce résultat montre la supériorité de cette nouvelle génération d'ondelette non seulement par rapport à l'ADO mais aussi par rapport aux informations classiques; notamment l'information économique, technique, retards (historique), et mesures de sentiment. En effet, l'APO effectue une meilleur analyse de la série temporelle (signal) dans toutes les sous-bandes fréquentielles possibles; ce que l'AOD ne fait pas. Par conséquent, l'information extraite par APO permet de mieux caractériser le signal d'origine et de retracer sa tendance via le canal basses-fréquences et ses changements brusques via son canal hautes-fréquences.

En somme, ce résultat est d'une importance extrême dans la mesure où il suggère que l'information intrinsèque représentée en temps-fréquence informe plus sur la direction du marché que les informations classiques. De ce fait, les travaux futurs devraient utiliser ce résultat dans leurs usages des systèmes hybrides sophistiqués.

RNA, SVM et fonction noyau

Contrairement à Tai et Cao (2001), Ajith et AuYeung (2003), et Kim (2003) qui ont conclu que le SVM est plus performant que le RNA dans le problème de la prédiction boursière, nos résultats obtenus des simulations montrent que le SVM ne performe pas mieux que les RNA. Ce constat est expliqué par le fait que le SVM est une machine statistique de classification qui suppose que les données sont identiquement

et indépendamment distribuées. Cette hypothèse est très forte. De plus, les données boursières sont très volatiles et ne suivent pas une distribution normale. Par contre, le RNA ne suppose pas que les données à modéliser soient identiquement et indépendamment distribuées et ne suppose pas que les données suivent une distribution normale. De ce fait, le RNA est théoriquement plus robuste pour la modélisation des données non normalement distribuées et très volatiles comme les données boursières. D'ailleurs, les résultats des simulations ont montré cela.

Par ailleurs, nous avons employé le SVM avec une fonction noyau polynomial pour deux raisons. La première est que la fonction noyau polynomial permet aux données séparées dans l'espace d'entrées d'avoir la même influence. Deuxièmement, la fonction noyau de type polynomial est plus adaptée à notre problème de classification (éventuelle hausse versus éventuelle baisse) par rapport à une fonction noyau de type fonction à base radiale. Cette dernière suppose que les données à modéliser suivent une distribution gaussienne. Or ce n'est pas le cas avec les données boursières, notamment la série temporelle de l'indice boursier S&P500. De plus, il faut bien déterminer ses paramètres; par exemple sa variance. En fait, nous avons utilisé la fonction noyau de type à base radiale en employant des paramètres estimés à partir de la distribution de la série temporelle du S&P500 à modéliser. Les résultats ne sont pas présentés dans notre thèse, mais nous avons trouvé qu'elle performait moins que le SVM avec la fonction noyau polynomial (d'ordre deux et plus) et aussi moins que les RNA.

Comparaison avec la littérature

Au niveau du taux de prédiction (classification), nous avons obtenus 72.2% et 70.9% en utilisant respectivement le RNA entraîné avec l'algorithme de Polak-Ribière et le SVM avec une fonction noyau polynomial d'ordre deux. Ces résultats sont très encourageants et même meilleurs que ceux qui ont été généralement obtenus dans la littérature comme on le voit dans le Tableau 5.1. De plus, selon Tsibouris et

Zeidenberg (1995), Baestaens et al. (1996), Walczak (2001), et Qian et Rasheed (2007) un taux de prédiction de 56% est satisfaisant pour la prédiction des mouvements du marché boursier vue le caractère volatile des séries temporelles boursières. De ce fait, notre approche qui utilise les RNA entraînés par les algorithmes numériques LM et Powell-Beale et qui utilisent de l'information fréquentielle obtenue par l'analyse en paquet d'ondelette est très prometteuse vue le taux de prédiction des mouvements futurs du S&P500 qui atteint 72.42%.

Au niveau de la littérature, le meilleur taux de prédiction obtenu de la tendance future du marché est de l'ordre de 73% par Huang, Nakamori, et Wang (2005) en utilisant un SVM pour la prédiction du mouvement du marché boursier japonais. Cependant, il faudrait noter que ce résultat a été obtenu sur une base d'échantillonnage hebdomadaire. De ce fait, la série temporelle boursière à modéliser est moins volatile que la série quotidienne comme la notre.

Tableau 5.1

Comparaison avec la littérature

Auteurs	outil	information	marché	performance
Fernandez-Rodriguez et al. (2000)	RNA	Historique	Madrid	54%-58%
Harvey et al (2000)	RNA	Économique	Korea	59%
Tay et Cao (2001)	SVM	Technique	CAC-40	53.76%
Halliday (2004)	RNARP	Historique	NYSE	55.57%
Doesken et al. (2005)	Mamdani Fuzzy Inference System	Technique	Microsoft	53.31%
Huang, Nakamori, et Wang (2005)	RNA (Elman)	Économique	Nikkai	69%
	SVM			73% (hebdomadaire)
Atsalakis et Valavanis (2009)	ANFIS	Historique	NYSE	68.33%
Azzini et al., (2010)	PSO	Techniques	S&P500	60%
	Negative Selection			63%
Cengiz et Herwartz (2011)	Régression partielle + modèle auto-régressif	Économique	DAX	57.9%
			FTSX	58.5%
			DOW	51.3%
Nyberg (2011)	Probit	Économique	S&P500	62.5%-63.4%
Approche de Wang et al., (2011) adoptée dans notre thèse	RNA	AOD	S&P500	56%
Notre approche (2011)	RNA (LM)	APO	S&P500	72.42%
	RNA (Powell-Beale)			

CHAPITRE VI

CONCLUSION

Le marché boursier est influencé par les conditions économiques, d'affaires, politiques, et aussi des états psychologiques des investisseurs. De ce fait, le marché boursier est dynamique, volatile et complexe. Le choix des variables prédictives détermine l'interprétation des résultats et le niveau de la performance des prévisions. Il existe en général quatre catégories d'information pour la prédiction boursière : information fondamentale, information technique, information comportementale, et information sur l'historique de la série des rendements boursiers.

Le premier objectif de cette thèse fut d'examiner la performance du réseau de neurones en termes de prédiction des hausses et des baisses de l'indice boursier S&P500 en fonction de la catégorie d'information disponible à l'investisseur qui constitue son entrée; à savoir l'historique récent (retards), information économique, indicateurs techniques, et mesures de sentiment du marché boursier. Pour chaque catégorie d'information nous avons utilisé les données qu'utilise généralement la littérature sur le marché boursier et qui sont rendues publiques.

Le deuxième objectif consistait à vérifier si l'agrégation de toute l'information disponible (économique, technique, retard, et mesures de sentiment des investisseurs) à l'investisseur puisse améliorer la prédiction correcte des hausses et des baisses du marché boursier.

Le troisième objectif était d'examiner le choix de la méthode numérique employée pour l'apprentissage du réseau sur la performance de prédiction. En particulier, nous avons comparé l'algorithme numérique standard et largement employé dans la littérature notamment la méthode de Levenberg-Marquardt (L-M) aux techniques numériques les plus avancées incluant le Quasi-Newton (BFGD), le gradient conjugué de type Powell-Beale, Fletcher-Reeves, et Polak-Ribière.

Le quatrième objectif fut de vérifier si l'information extraite par l'analyse en paquet d'ondelette permet d'améliorer la qualité de la prédiction de la tendance future du marché S&P500.

Finalement, les statistiques taux de classification correct, sensibilité et spécificité ont été utilisées pour évaluer la performance du réseau de neurones artificiels (RNA). Nous avons également comparé sa performance aux machines à supports de vecteurs (SVM).

Nous avons utilisé les techniques statistiques pour prétraiter les données et d'y extraire les informations pertinentes à la modélisation par réseaux de neurones et machines à supports de vecteurs. En particulier, les méthodes statistiques étaient employées pour rendre les données stationnaires, modéliser le processus des rendements et extraire sa volatilité, et finalement pour déterminer les variables prédictives.

Finalement, nous avons tiré les conclusions relatives aux objectifs de la thèse comme suit : La première conclusion est que les informations de types historiques et psychologiques sont les pertinentes pour la prédiction de la tendance (hausse et baisses) future du S&P500. La deuxième conclusion est que la combinaison des différentes catégories d'information (économique, technique, historique, sentiment) n'améliore pas la précision de la prédiction de la tendance future du marché S&P500. La troisième conclusion est qu'en général qu'il n'y a pas de différences majeures

entre les performances des algorithmes numériques étudiés dans le cadre de cette thèse dans la prédiction de la tendance future du marché boursier S&P500. Cependant, on pourrait dire que les algorithmes basés sur le gradient conjugué (exemple : Polak-Ribière) performant en général mieux que BFGS et LM.

La quatrième conclusion est que l'information fréquentielle extraite par l'analyse multi-résolution par paquet d'ondelettes (APO) permet largement d'améliorer la performance prédictive du RNA par rapport aux autres catégories d'information traditionnellement utilisées dans la littérature; et ce quelque soit l'algorithme numérique utilisé pour l'entraînement. De plus, l'usage de l'information extraite par APO permet d'obtenir une meilleure prédiction de la tendance future du marché S&P500 que l'information extraite par analyse en ondelette standard qui est utilisée dans la littérature. En effet, sur la base de la même fréquence quotidienne des données, nous avons obtenu une détection correcte de la tendance future du marché égale à 72.42% contre 54% à 59% obtenue dans la littérature. En outre, l'approche par l'APO est beaucoup plus simple à utiliser contrairement aux approches traditionnelles qui requièrent beaucoup de prétraitements statistiques.

La cinquième conclusion est que le RNA performe généralement mieux que le SVM selon nos données du marché S&P; quoique les différences soient minimes. Il est également important de noter que le fait d'augmenter le nombre de couches cachées n'améliore pas le taux de classification des RNA. Au contre, il le fait diminuer. De ce fait, une seule couche cachée permet d'obtenir de meilleurs résultats dans notre cas.

Par ailleurs, la performance des RNA varie avec le nombre de neurones dans la couche cachée. Il faudrait rappeler le lecteur qu'à ce niveau le nombre de neurones a été varié seulement dans le cas d'une seule couche cachée.

La sixième conclusion est que pour n'importe quelle information conventionnelle utilisée par le classifieur; ce dernier ne détecte pas correctement les tendances baissières dans la plupart des cas. Cependant, la spécificité des classifieurs s'améliore clairement avec l'usage de l'information obtenue par l'analyse en paquet d'ondelette.

La septième conclusion est que les résultats obtenus par validation croisée confirment les résultats obtenus sans validation croisée. Cependant, il faudrait être prudent avec les résultats de la validation croisée parce que les données en question sont des séries temporelles et en faisant de la validation croisée la dynamique de l'information contenue dans les données a été interrompue. Nous laissons le soin au lecteur d'en évaluer la pertinence.

La huitième conclusion est que la prédiction du marché boursier est définitivement une tâche très difficile à cause du fait que les séries temporelles qui caractérisent les mouvements du marché sont complexes et difficiles à modéliser. D'ailleurs, les taux de classification sont généralement entre 51% et 58%. Dans le meilleur cas, le taux de classification atteint 72.43% : notre approche basée sur la transformée en paquet d'ondelette pour extraire de l'information intrinsèque de la série temporelle en question. Malgré cela, ce taux de classification reste bien bas comparativement à d'autres problèmes de classification en science et en génie où les taux de classification sont très élevés. On peut citer par exemple le problème de classification des images biomédicales dans le cadre du diagnostic médical automatisé où les taux de classification sont supérieurs à 90%.

Finalement, nous voudrions apporter quelques critiques à notre travail. Premièrement, nous n'avons pas utilisé d'autres indices boursiers (américains et internationaux) et d'autres périodes pour pouvoir généraliser nos conclusions. Deuxièmement, nous n'avons pas testé nos systèmes sur différentes sous-périodes qui composent la période d'études afin de vérifier la stabilité des résultats. Cependant,

nos résultats sont encourageants dans la mesure où la période de test est caractérisée par une haute volatilité du marché.

Troisièmement, nous n'avons pas considéré différentes ondelettes mères et différents niveau d'analyse multi-résolution pour examiner leurs effets sur la performance de prévision. Quatrièmement, nous n'avons pas mesuré la performance des prévisions en termes de rendements d'un portefeuille hypothétique.

La principale conclusion dans notre thèse est que l'extraction de l'information par paquet d'ondelettes et son traitement par un réseau de neurones s'avère être une orientation prometteuse pour les recherches futures en termes de modélisation des séries temporelles boursières, notamment la prédiction de la tendance du marché boursier. Pour les travaux futurs nous proposons, dans l'immédiat, d'améliorer notre meilleur système obtenu en examinant l'effet du choix de l'ondelette mère.

À moyen terme, nous envisageons de poursuivre la recherche en considérant les quatre critiques mentionnées précédemment. Finalement et à long terme, nous recommandons de répartir l'effort de la recherche sur différents axes. Par exemple, documenter d'autres mesures psychologiques qui peuvent être fournis par la finance comportementale ou la psychologie de la décision et d'investissement. Également, se pencher sur la littérature en finance pour vérifier si d'autres variables explicatives ont été proposées. Sélectionner les variables par le test de causalité de Granger et examiner l'effet marginal de chacune sur le système prédictif intelligent et aussi étudier l'effet marginal du nombre de variables prédictives. On peut également considérer les techniques de sélection des caractéristiques souvent utilisées en reconnaissance de forme (exemple, iterative approach et wrapper approach) pour sélectionner les meilleurs variables - parmi l'ensemble des variables sélectionnées par le test de Granger - qui vont constituer les entrées aux systèmes prédictifs. Ce type d'examen est aussi utile pour la construction d'un indice d'information synthétique, au besoin, de façon intelligente. Un autre axe de recherche qui est très intéressant

serait de décomposer la série temporelle des rendements boursiers en différents régimes (cycles) et puis concevoir différents systèmes intelligents et hybrides qui soient où chaque sous-système est spécialisé dans la modélisation de la série dans un segment (régime) donné.

Plus important encore et comme nous l'avons déjà mentionné dans le chapitre qui passe en revue les travaux de la littérature, nous prévoyons d'employer les réseaux de neurones spatio-temporels pour modéliser un titre boursier côté simultanément dans plusieurs places boursières mondiales et de modéliser l'interaction de plusieurs marchés boursiers internationaux comme un seul système dynamique. Nous envisageons également de procéder à une analyse chaotique à court terme et d'en tester les implications sur les prévisions à court terme dans le cadre d'une gestion simulée d'un portefeuille.

Un autre axe de recherche consiste à reconsidérer les simulations sur la période 2007-2011 caractérisé par une forte volatilité à cause de la crise financière internationale afin d'examiner la validité des conclusions tirées pour la période 2003-2007. Finalement, il est également intéressant d'utiliser des techniques de computation évolutive (algorithmes génétiques, optimisation par essaim, etc) pour optimiser de manière automatique la topologie et le nombre de neurones dans la couche cachée du RNA. Cette approche permettrait d'élargir le champ de la recherche pour trouver l'optimal si il existe et d'éviter l'approche essai-erreur.

Finalement, l'approche proposée dans cette thèse peut être appliquée à toutes données financières, économiques, ou d'affaires (ventes de produits, consommation d'un bien). Par exemple, si l'on veut prédire les ventes dans un secteur d'activité donné on pourrait consituter initialement un ensemble de données composées d'information sur le secteur, sur l'industrie, les conditions financières et économiques et sur le comportement du consommateur. Ensuite, le test de Granger peut être appliqué pour identifier les variables qui sont significativement associées aux variations aux ventes

du produit en question. Finalement, les RNA et SVM seraient utilisés pour fournir les prédictions.

En somme, il reste beaucoup à faire en modélisation des séries temporelles boursières pour fins de prévisions.

APPENDICE A

LE CAPITAL ASSET PRICING MODEL

Il existe plusieurs versions du CAPM qui sont très similaires. Nous allons brièvement décrire la version de Sharpe (1964) et Linter (1966). Le CAPM peut être exprimé de la forme suivante :

$$E(R_{i,t}) = R_f + [E(R_{m,t}) - R_{f,t}] \beta_{i,m} \quad (6.1)$$

Où $E(R_{i,t})$ est le rendement anticipé de l'actif i , R_f est le rendement sans risque (exemple : bond du trésor), $E(R_{m,t})$ est le rendement anticipé du marché boursier, et $\beta_{i,m}$ est un paramètre qui représente la mesure du risque du marché (ou la sensibilité du marché appelé aussi risque systématique). Empiriquement, le CAPM est simplement représenté sous la forme de la régression linéaire suivante :

$$R_{i,t} = \alpha_i + \beta_i R_{m,t} + \varepsilon_{i,t} \quad (6.2)$$

$$\varepsilon \sim N(0, \sigma^2) \quad (6.3)$$

La valeur de β_i est donnée par :

$$\beta_i = \frac{Cov(R_{i,t}, R_{m,t})}{Var(R_{i,t})} \quad (6.4)$$

APPENDICE B

MODÉLISATION DU PROCESSUS ARMA ET EGARCH

Les modélisations des processus ARMA et EGARCH supposent au préalable que les variations $R(t)$ de l'indice boursier $I(t)$ suivent un processus ARMA(p, q) - appelée l'équation de la moyenne – suivant (Box et Jenkins, 1970):

$$R_t = \alpha_s + \sum_{i=1}^p \rho_{i-1} R_{t-1} + \sum_{i=1}^q \phi_{i-1} \varepsilon_{t-1} + \varepsilon_t \quad (6.5)$$

$$\varepsilon_t \sim GED(0, h_t) \quad (6.6)$$

Où ρ et ϕ sont les coefficients des termes autorégressifs et termes d'erreurs ε respectivement. Le terme d'erreur est supposé suivre une distribution exponentielle généralisée (Generalized Exponential Distribution, GED) dont la moyenne est zéro et sa variance $h_{t,t}$ est hétérosédastique : dépendante du temps. La GED est calculé selon la formule suivante:

$$f(\varepsilon, d) = \frac{d \Gamma\left(\frac{3}{d}\right)^{1/2}}{2 \Gamma\left(\frac{1}{d}\right)^{3/2}} \exp \left[-|\varepsilon|^d \left(\frac{\Gamma\left(\frac{3}{d}\right)}{\Gamma\left(\frac{1}{d}\right)} \right)^{d/2} \right] \quad (6.7)$$

Où, $-\infty < \varepsilon < +\infty$, d est un paramètre ($d > 0$), et Γ est la fonction Gamma définit par :

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt, x \geq 0 \quad (6.8)$$

Les ordres p et q de l'équation de la moyenne ci-haut sont déterminés en minimisant les critères d'information d'Akaike (Akaike information criterion, AIC) et de Schwarz (Schwarz information criterion, SIC) calculés comme suit :

$$AIC = -2\ell/T + 2k/T \quad (6.9)$$

$$SIC = -2\ell/T + (k \log T)/T \quad (6.10)$$

$$\ell = \frac{-T}{2} (1 + \log(2\pi) + \log(\hat{\varepsilon}'_s \hat{\varepsilon}_s / T)) \quad (6.11)$$

Où, k et T sont respectivement le nombre de coefficients et la taille de l'échantillon utilisé pour l'estimation de l'équation de la moyenne et ℓ est la fonction de vraisemblance. Pour s'assurer que le processus ARMA(p, q) est correctement spécifié, le test d'auto corrélation des résidus de Breusch-Godfrey⁴⁵ (1981) est employé. Ensuite, la variance conditionnelle $h_{i,t}$ est modelée selon le processus EGARCH (Exponential GARCH; Nelson, 1991). Le processus EGARCH(m, n) est donné par l'équation de la variance suivante :

$$\log(h_t) = w + \sum_{i=1}^m \Phi_j \left[\frac{|\varepsilon_{t-i}|}{\sqrt{h_{t-i}}} - \sqrt{2/\pi} \right] + \gamma \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \sum_{j=1}^n \beta_j \log h_{t-j} + \eta_t \quad (6.12)$$

Où, w , Φ_j , γ , et β_j sont des paramètres constants. Le processus EGARCH n'impose pas de restrictions sur ses paramètres à estimer. La variance $h_{i,t}$ est définit comme

⁴⁵Voir annexe G.

une fonction asymétrique des bruits retardés ε_{t-1} . Les ordres m et n de l'équation de variance sont également déterminés en minimisant les critères d'information statistiques mentionnés en haut. Ensuite, le test ARCH LM⁴⁶ introduit par Engle (1982) est employé pour s'assurer que le processus EGARCH est correctement identifié.

Finalement, la série temporelle de la variance conditionnelle estimée est extraite. En effet, l'objectif est d'estimer les séries des variances conditionnelles des volatilités des rendements et du volume.

Selon les critères d'information des estimations montrés dans les Tableaux B.1 et B.2, le processus ARMA(2,0)⁴⁷ est retenu pour la modélisation de la série temporelle $R(t)$ des variations de l'indice boursier. Il est important d'identifier le bon processus car la qualité des estimations statistiques ultérieures et des prévisions en dépend énormément.

Les résultats du test d'auto corrélation des résidus de Breusch-Godfrey (1981) montrent que le processus identifié ARMA(2,0) est correctement spécifié⁴⁸. Finalement, les estimations du processus ARMA(2,0) avec modélisation de la variance conditionnelle par EGARCH(1,1) sont présentées dans le Tableau B.3, et les résultats du test ARCH LM introduit par Engle (1982) montrent qu'il n'y a pas d'effet ARCH et donc le processus EGARCH(1,1) est correctement identifié⁴⁹. De même, le bon processus (Tableau B.4) pour la série du volume des transactions est

⁴⁶Voir annexe H.

⁴⁷Il faudrait noter que malgré le fait que les coefficients du processus ARMA(1,1) sont hautement significatifs et que les coefficients du processus ARMA(2,0) ne sont pas du tout significatifs, ce dernier est retenu car les critères d'information le suggèrent.

⁴⁸Voir annexe G.

⁴⁹Voir annexe H.

ARMA(4,4)⁵⁰ et le processus de sa variance conditionnelle est EGARCH(1,1). En somme, les bons processus sont identifiés⁵¹ (tous les paramètres sont statistiquement efficaces et les séries des volatilités des rendements et du volume sont extraites).

Tableau B.1

Résultats de l'estimation du processus ARMA(1,1)

Coefficient	Valeurs	Erreur Standard	Statistique t	Probabilité
α	-3.02E-05	0.000107	-0.283347	0.7769
$\rho(t-1)$	0.76663	0.235453	3.256006	0.0012
$\phi(t-1)$	0.80175	0.220986	-3.628092	0.0003
ℓ	6242.70			
	9			
AIC	7.74638			
	0			
SC	7.73635			
	3			

⁵⁰Pour alléger le texte, seuls les estimés et statistiques du bon processus identifié ARMA(4,4) sont figurent dans cette thèse.

⁵¹On peut vérifier le bon choix du processus ARMA pour toutes les séries en examinant également les fonctions d'auto corrélation, d'auto corrélation partielle et la statistique Q . La définition des fonctions et des statistiques ainsi que les résultats sont donnés en annexe E.

Tableau B.2

Résultats de l'estimation du processus AR(2)

Coefficient	Valeurs	Erreur	Statistique t	Probabilité
		Standard		
α	-2.33E-05	0.000116	-0.200351	0.8412
$\rho(t-1)$	-0.027907	0.032942	-0.847162	0.3970
$\rho(t-2)$	-0.046384	0.035827	-1.294669	0.1956
ℓ	6241.140			
AIC	-7.749242			
SC	-7.739210			

Tableau B.3

Résultats de l'estimation du processus ARMA(2,0)-EGARCH(1,1)

Coefficient	Valeurs	Erreur	Statistique t	Probabilité
		Standard		
α	-6.32E-05	8.89E-05	-0.711364	0.4769
$\rho(t-1)$	-0.038564	0.024696	-1.561544	0.1184
$\rho(t-2)$	-0.029415	0.026372	-1.115404	0.2647
w	-0.173785	0.032612	-5.328959	0.0000
ϕ	0.064704	0.018282	3.539297	0.0004
γ	-0.097890	0.013997	-6.993554	0.0000
β	0.988798	0.002420	408.5204	0.0000

Tableau B.4

Résultats de l'estimation du processus ARMA(4,4)-EGARCH(1,1)

Coefficient	Valeurs	Erreur Standard	Statistique t	Probabilité
α	-0.000171	0.000659	-0.259173	0.7955
$\rho(t-1)$	0.197656	0.205852	0.960185	0.3370
$\rho(t-2)$	-0.549624	0.160015	-3.434841	0.0006
$\rho(t-3)$	-0.464399	0.227544	-2.040916	0.0413
$\rho(t-4)$	0.248936	0.056921	4.373392	0.0000
$\phi(t-1)$	-0.642192	0.208032	-3.086987	0.0020
$\phi(t-2)$	0.466162	0.259973	1.793120	0.0730
$\phi(t-3)$	0.135116	0.269255	0.501812	0.6158
$\phi(t-4)$	-0.487043	0.147040	-3.312325	0.0009
w	-1.605181	0.195256	-8.220882	0.0000
Φ	0.514929	0.050638	10.16886	0.0000
γ	-0.031775	0.032946	-0.964431	0.3348
β	0.680218	0.046668	14.57552	0.0000
ℓ	1187.658			
AIC	-1.168400			
SC	-1.129338			

APPENDICE C

LE TEST STATISTIQUE DE LA STATIONARITÉ

Le test de Dickey-Fuller (1979) augmenté (Augmented Dickey-Fuller Test) construit une correction paramétrique pour les corrélations d'ordre supérieur en supposant que la série à modéliser y (les séries de rendements dans notre cas) suit un processus $AR(p)$ où p est l'ordre du régime autorégressif.

$$\Delta y_t = \alpha y_{t-1} + x'_t \delta + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \dots + \beta_p \Delta y_{t-p} + v_t \quad (6.13)$$

Où x est un vecteur optionnel contenant le terme de constante et le trend, et α , β , et δ sont des paramètres à estimer, et v est le terme d'erreur. Ce test est utilisé pour tester la stationnarité de la série y selon les hypothèses nulle et alternative définies comme suit :

$$\begin{aligned} H_0 : \alpha &= 0 \\ H_1 : \alpha &< 0 \end{aligned} \quad (6.14)$$

Le test est évalué en utilisant le *t-ratio* conventionnel pour α :

$$t_\alpha = \frac{\hat{\alpha}}{se(\hat{\alpha})} \quad (6.15)$$

Où, $\hat{\alpha}$ est l'estimé de α , et $se(\hat{\alpha})$ est l'écart type de $\hat{\alpha}$.

Les résultats des tests de stationnarité sont donnés dans le Tableau C.1 qui montre que les variables R , ICI , VIX , $Put\ Call\ Ratio$, et la volatilité du volume sont

stationnaires. Les variables prédictives qui ne sont pas stationnaires vont être transformées en différences premières pour les rendre stationnaires⁵².

Le Tableau C.2 montre que les nouvelles séries transformées - des différences premières des séries non stationnaires - sont rendues stationnaires.

Tableau C.1

Résultats du test ADF de stationarité

Variables	Statistique-t	Probabilité
Rendement	-46.74920	0.0001
DAAA	-1.867724	0.3479
DBAA	-1.641623	0.4610
DTB3	-1.214688	0.6703
DTB6	-1.264576	0.6480
DFEDTAR	-1.089329	0.7221
DFF	-1.240942	0.6587
DEXCAUS	0.240843	0.9751
DEXJPUS	-2.282044	0.1780
DEXSZUS	-0.758393	0.8298
DEXUSEU	0.170917	0.9708
DEXUSUK	-0.482753	0.8921
DTWEXB	0.738637	0.9930
DTWEXM	0.182775	0.9715
ICI	-2.595093	0.0441
VIX	-44.91632	0.0001
Put Call Ratio	-5.042192	0.0000
Volume	-1.526088	0.5203
Volatilité de R(t)	-31.53885	0.0000

⁵²Par exemple, si une série temporelle $x(t)$ est non stationnaire alors la série temporelle définie comme $x' = x(t) - x(t-1)$ peut être stationnaire. Si $x'(t)$ est stationnaire on dit que la série $x(t)$ est intégrée d'ordre 1. Si la nouvelle série $x'(t)$ n'est pas stationnaire alors on définit une autre série temporelle $x'' = x'(t) - x'(t-1)$ qui est généralement stationnaire est on dit que la série $x(t)$ est intégrée d'ordre 2.

Tableau C.2

Résultats du test de stationarité pour les variables transformées

Variables	Statistique-t	Probabilité
DAAA'	-44.16908	0.0001
DBAA'	-44.23027	0.0001
DTB3'	-27.79268	0.0000
DTB6'	-17.10352	0.0000
FEDTAR'	-44.84593	0.0001
DFF'	-18.86962	0.0000
DEXCAUS'	-45.11693	0.0001
DEXJPUS'	-46.42256	0.0001

APPENDICE D

LE TEST DE CAUSALITÉ DE GRANGER

Nous avons employé le test de causalité de Granger (1969) pour identifier les entrées (variables prédictives) qui ont un effet causal statistiquement significatif - à un seuil de confiance 5% - sur la série temporelle des variations $R(t)$. Le test est basé sur des régressions bi-variées comme suit :

$$\begin{aligned} y_t &= \beta_{y,0} + \beta_{y,1}y_{t-1} + \dots + \beta_{y,k}y_{t-k} + \delta_{x,1}x_{t-1} + \dots + \delta_{x,k}x_{t-k} + \eta_t \\ x_t &= \beta_{x,0} + \beta_{x,1}x_{t-1} + \dots + \beta_{x,k}x_{t-k} + \delta_{y,1}y_{t-1} + \dots + \delta_{y,k}y_{t-k} + \nu_t \end{aligned} \quad (6.16)$$

Où η et ν représentent des bruits gaussiens. Ensuite, les statistiques F sont calculées comme une statistique de Wald (Davidson et MacKinnon, 1993) pour tester l'hypothèse nulle conjointe :

$$\delta_1 = \delta_2 = \dots = \delta_k = 0 \quad (6.17)$$

Les statistiques F permettent de tester si les coefficients des variables retardées x sont statistiquement significatifs pour expliquer la variable dépendante y . Dans notre thèse, le nombre de retards k est arbitrairement mis égale à 5. Le Tableau D.1⁵³

⁵³ Il faut noter que les variables SDFF, SDEXSZUS, SDTWEXB, et SDXJPUS sont déjà rendu stationnaires avant de performer le test de causalité de Granger.

montre seulement⁵⁴ les résultats statistiquement significatifs obtenus. Les variables qui y figurent sont les seules qui causent une variation significative dans la variation $R(t)$ de l'indice boursier, ou R tout court. Ces variables sont les entrées à chaque classificateur. Les nouvelles séries transformées et stationnaires sont présentées dans la Figure D.1.

Tableau D.1

Détails des résultats du test de causalité de Granger

Hypothèse nulle	Statistique-F	Probabilité
SDFF ne cause pas R	3.23919	0.00645
SDEXSZUS ne cause pas R	2.50074	0.02884
SDTWEXB ne cause pas R	2.16523	0.04953
SDXJPUS ne cause pas R	2.36904	0.03735
X4	3.5037	0.0038
EMV	3.0764	0.00918
Volatilité de R ne cause pas R		
R	2.57838	0.01628
ICI ne cause pas R	3.29216	0.00578

⁵⁴Comme le test de causalité est un test bidirectionnel et que nous avons 19 variables, donc le nombre total de tests à effectuer est de l'ordre $19 \times 2 = 38$ tests. Ainsi, nous avons jugé qu'il est pertinent de ne rapporter que les résultats des tests dont les statistiques F sont significatives.

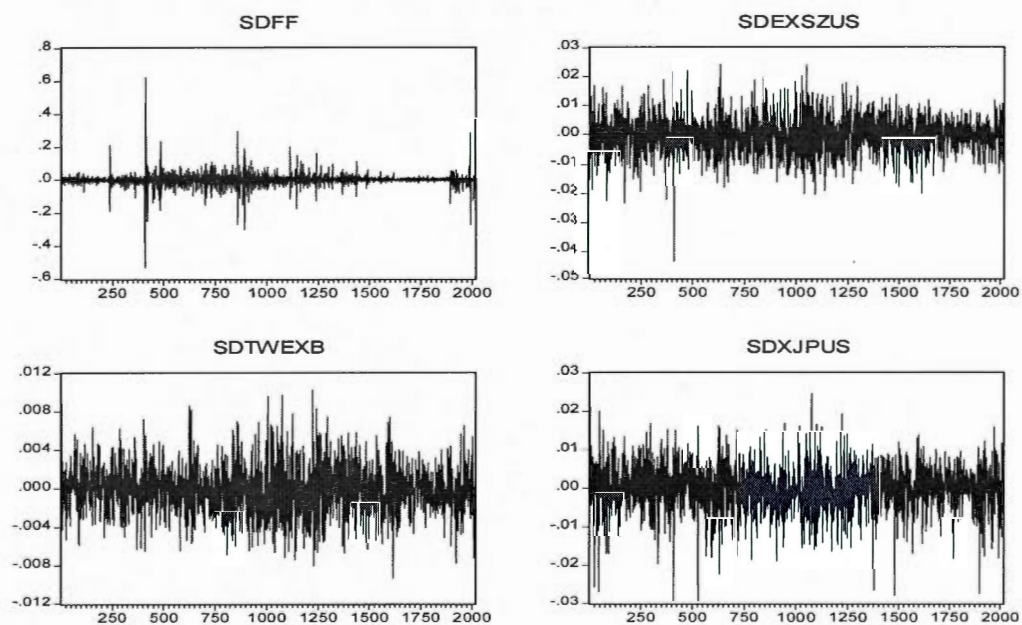


Figure D.1 Séries transformées et stationnaires

APPENDICE E

FONCTIONS D'AUTOCORRÉLATION ET D'AUTOCORRÉLATION
PARTIELLE

La fonction d'auto-corrélation (AC) (Davidson et MacKinnon, 1993) de la série temporelle $R(t)$ des variations de l'indice boursier à k retards est donnée par l'équation suivante :

$$\tau_k = \frac{\sum_{t=k+1}^T (R(t) - \bar{R})(R(t-k) - \bar{R})}{\sum_{t=1}^T (R(t) - \bar{R})^2} \quad (6.18)$$

Où \bar{R} est la moyenne empirique de $R(t)$.

La fonction d'auto corrélation (Davidson et MacKinnon, 1993) à k retard est obtenue en estimant la régression suivante :

$$R(t) = \beta_0 + \beta_1 R(t-1) + \dots + \beta_{k-1} R(t-(k-1)) + \phi_k R(t-k) + \varepsilon(t) \quad (6.19)$$

Ensuite, les coefficients ϕ sont extraits. Cette procédure est plus efficiente, précise et simple que la méthode de Box et Jenkins (1976) qui repose essentiellement sur une estimation de façon récursive qui est coûteuse sur le plan computationnel et aussi n'est pas efficiente. La statistique Q (Q-statistic) (Ljung et Box, 1978) est définit comme :

$$Q_k = T(T+2) \sum_k^T \rho_j^2 (T-j)^{-1} \quad (6.20)$$

Où ρ_j est le j -ième retard d'auto corrélation, k est les nombre d'auto corrélations, et T est la taille de l'échantillon. L'hypothèse nulle est que : $H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0$. Le rejet de l'hypothèse nulle indique la présence d'auto corrélation dans la série temporelle. Le test est largement décrit dans Ljung et Box (1978).

Les résultats de l'examen de la fonction d'auto corrélation confirment que la série temporelle des rendements suit un processus ARMA(2,0). Ils sont donnés dans le Tableau 13. Les trois statistiques à savoir la fonction d'auto corrélation (AC), d'auto corrélation partielle (ACP) et la statistique Q de la série temporelle des rendements sont données dans le Tableau E.1. Les résultats à 5 retards du Tableau E.1 confirment le bon choix du processus ARMA(2,0) de la série temporelle des variations $R(t)$ de l'indice boursier S&P500.

Tableau E.1

Résultats de la fonction d'autocorrélation

k	AC	AC Partielle	Statistique-Q	Probabilité
1	-0.042	-0.042	3.5673	0.059
2	-0.035	-0.037	6.0458	0.049
3	0.003	0.000	6.0603	0.109
4	-0.001	-0.002	6.0614	0.195
5	-0.025	-0.026	7.3722	0.194

APPENDICE F

LA MATRICE HAC DE NEWEY-WEST

L'estimateur de matrice de covariance de Newey-West (1987) est donné par :

$$\hat{\Sigma}_{NW} = \frac{T}{T-k} (X'X)^{-1} \hat{\Omega} (X'X)^{-1} \quad (6.21)$$

Où,

$$\hat{\Omega} = \frac{T}{T-k} \left\{ \sum_{t=1}^T \varepsilon_t^2 x_t x_t' + \sum_{v=1}^q \left(\left(1 - \frac{v}{q+1} \right) \sum_{t=v+1}^T (x_t \varepsilon_t \varepsilon_{t-v} x_{t-v}' + x_{t-v} \varepsilon_{t-v} \varepsilon_t x_t') \right) \right\} \quad (6.22)$$

Où, q est le paramètre qui représente le nombre d'auto-corrélations utilisées pour l'évaluation de la dynamique des résidus ε_t des moindres carrés. Le paramètre v ($v=0;1;2\dots p$) indique l'ordre de la matrice d'auto covariance. En suivant la suggestion de Newey et West (1987), le paramètre q est défini comme :

$$q = \text{floor} \left(4 \left(\frac{T}{100} \right)^{2/9} \right) \quad (6.23)$$

APPENDICE G

TESTS D'AUTOCORRELATION DES RÉSIDUS

Le test de Breusch-Godfrey (1981) est un test basé sur la statistique LM (Lagrange Multiplier⁵⁵) pour tester l'auto corrélation dans le terme d'erreur. Le test présente plusieurs avantages. Il est utilisable pour des ordres supérieures du terme d'erreur dans le processus ARMA. En deuxième lieu, il est applicable qu'il y a ou non des variables dépendantes retardées comme variables explicatives. La statistique du test est calculée par une régression auxiliaire selon deux étapes. L'équation d'origine est estimée dans sa forme générale comme suit⁵⁶ :

$$y_t = X_t \beta + \varepsilon_t \quad (6.24)$$

Où, β et ε sont les matrices des coefficients et des erreurs estimés respectivement à l'instant t . ensuite, l'équation auxiliaire des résidus ε sur les variables exogènes (explicatives) et les résidus retardés jusqu'à ordre p ; cette équation est estimée comme :

$$\varepsilon_t = X_t \gamma + \left(\sum_{s=1}^p \alpha_s \varepsilon_{t-s} \right) + v_t \quad (6.25)$$

⁵⁵Multiplicateur de Lagrange.

⁵⁶Dans notre cas, l'équation d'origine à estimer est:

$$R_t = C + \sum_{i=1}^p a_i R_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t.$$

Où ν est le terme d'erreur. L'hypothèse nulle est qu'il n'y a pas d'auto corrélation dans les résidus ε jusqu'à l'ordre p , et la statistique LM est calculée comme :

$$LM = (n - p)R^2 \quad (6.26)$$

Où n est le nombre d'observations et R^2 (le R carré) est la statistique calculée à partir de la régression auxiliaire à tester. La statistique LM est asymptotiquement distribuée comme une $\chi^2(p)$. Les résultats sont donnés dans les Tableaux G.1 et G.2 pour les séries temporelles $R(t)$ et sa volatilité $V(t)$.

Tableau G.1

Résultat du test Breusch-Godfrey LM pour AR(2)

F-statistic	1.399125	Prob. F(4,2005)	0.231804
Obs*R-squared	5.600407	Prob. Chi-Square(4)	0.231044

Tableau G.2

Résultat du test Breusch-Godfrey LM pour ARMA(4,4)

F-statistic	0.676074	Prob. F(4,1996)	0.608546
Obs*R-squared	2.651212	Prob. Chi-Square(4)	0.617778

APPENDICE H

LE TEST D'HETEROSKEDASTICITÉ ARCH-LM

L'effet d'hétéroskedasticité autorégressive conditionnelle (autoregressive conditional heteroskedasticity, ARCH) dans les résidus ε de la régression de type moindre carré ordinaires peut conduire à un manque d'efficience dans l'estimation des coefficients. Engle (1982) a proposé un test ARCH LM pour tester l'hypothèse nulle qu'il y a pas d'effet ARCH jusqu'à ordre q dans les résidus en estimant l'équation auxiliaire suivante :

$$\varepsilon_t^2 = \beta_0 + \left(\sum_{s=1}^q \beta_s \varepsilon_{t-s} \right) + \eta_t \quad (6.27)$$

Où, ε_t est le terme d'erreur de la régression d'origine (exemple : $y_t = X_t\beta + \varepsilon_t$). La statistique LM d'Engle est calculée comme suit :

$$LM = (n - q)R^2 \quad (6.28)$$

Où n est le nombre d'observations et R^2 est la statistique calculée à partir de la régression auxiliaire à tester. La statistique LM est asymptotiquement distribuée comme une $\chi^2(p)$. Les résultats sont donnés dans les Tableaux H.1 et H.2.

Tableau 1 Résultat du test ARCH LM pour ARMA(2,0)-EGARCH(1,1)

Tableau H.1

Résultat du test ARCH LM pour ARMA(2,0)-EGARCH(1,1)

F-statistic	1.054891	Prob. F(4,1601)	0.377480
Obs*R-squared	4.221615	Prob. Chi-Square(4)	0.376843

Tableau H.2

Résultat du test ARCH LM pour ARMA(4,4)-EGARCH(1,1)

F-statistic	0.742298	Prob. F(4,2000)	0.563115
Obs*R-squared	2.972202	Prob. Chi-Square(4)	0.562488

BIBLIOGRAPHIE

Ajith A., AuYeung A. (2003). Integrating Ensemble of Intelligent Systems for Modeling Stock Indices. IWANN 2003, LNCS 2687, 774-781.

Ajith A., Baikunth N., Mahanti P.K. (2001). Hybrid Intelligent Systems for Stock Market Analysis. LNCS 2074, 337-345.

Ajith A., Baikunth N., Mahanti P.K. (2003). Hybrid intelligent systems for stock market analysis. In Proceedings of International Conference on Computational Science.

Alligood K.T., Sauer T., Yorke, J.A. (1997). Chaos: an introduction to dynamical systems. Springer-Verlag New York.

Andersen T.G., Bollerslev T., Christoffersen P., Diebold F.X. (2006). Volatility and Correlation Forecasting. In: Elliott G, Granger CW J, Timmermann A (eds.), Handbook of Economic Forecasting, Amsterdam: North-Holland, 778-878.

Andersen T.G., Bollerslev T., Diebold F.X., Labys P. (2003). Modeling and Forecasting Realized Volatility. *Econometrica*, 71, 529-625.

Andersen T.G., Bollerslev T., Lange S. (1999). Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast Horizon. *Journal of Empirical Finance*, 6, 457-477.

Ang A., Hodrick R.J., Xing Y., Zhang X. (2005). The Cross-Section of Volatility and Expected Returns. *Journal of Finance*, 61, 259-300.

Armano G., Marchesi M., Murru A. (2004). A hybrid genetic-neural architecture for stock indexes forecasting. *Information Sciences*, 170 (1), 3-33.

Atiya A. (1992). Nonlinear System Identification using Spatiotemporal Neural Networks.

International Joint Conference on Neural Networks, 2, 504-509.

Atsalakis G.S., Valavanis K.P. (2009) Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert Systems with Applications*, 36, 5932-5941.

Atsalakis G.S., Valavanis K.P. (2009). Forecasting stock market short-term trends using a neuro-fuzzy based methodology. *Expert Systems with Applications* 36, 10696-10707.

Aussem A., Campbell J., Murthagh F. (1998). Wavelet-based feature extraction and decomposition strategies for financial forecasting, *Journal of Computational Intelligence in Finance*, 6 (2), 5-12.

Avramov D., Kosowski R., Naik N.Y, Teo M. (2011). Hedge Funds, Managerial Skill, and Macroeconomic Variables. *Journal of Financial Economics*, 99, 672-692.

Azzini A., De Felice M., Tettamanzi A.G.B. (2010). A Study of Nature-Inspired Methods for Financial Trend Reversal Detection. *Evo Applications, Part II, LNCS* 6025, 161-170.

Baba N., Kozaki M. (1992). An intelligent forecasting system of stock price using neural networks. *Proceedings of the IEEE International Joint Conference on Neural Networks*, 371-377.

Baek J., Cho S. (2002). Time to jump. Long rising pattern detection in KOSPI 200 future using an auto-associative neural network. *Lecture Notes in Computer Science* (2412). Springer.

Baestaens D.J.E, Van Den Bergh W.M., Vaudrey H. (1996). Market inefficiencies, technical trading and neural networks. In: Dunis C (ed) forecasting financial markets. Financial economics and quantitative analysis. John Wiley & Sons, Chichester, England, 254-260.

Baker G.L. (1996). Chaos, Scattering and Statistical Mechanics. Cambridge University Press.

Baker M., Wurgler J. (2004). A catering theory of dividends. *Journal of Finance*, 59, 1125-1165.

Baker M., Wurgler J. (2000). The equity share in new issues and aggregate stock returns. *Journal of Finance*, 55, 2219-2257.

Baker M., Wurgler J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, American Economic Association, 21 (2), 129-152.

Bandopadhyaya A., Jones A.L. (2008). Measures of Investor Sentiment: A Comparative Analysis Put-Call Ratio Vs. Volatility Index. *Journal of Business & Economics Research*, 6 (8), 27-34.

Barber B.M., Odean T. (2000). Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors. *Journal of Finance*, 55 (2), 773-806.

Barber M., Odean T., Zhu N. (2009). Do retail trades move markets?. *Review of Financial Studies*, 22, 151-186.

Barberis N., Shleifer A., Vishny R. (1998). A Model of Investor Sentiment. *Journal of Financial Economics*, 49 (3), 307-343.

Bautista, C. C. (2001). Predicting the Philippine Stock Price Index using artificial neural networks. UPCBA Discussion Paper No. 0107.

Benjamin G., David D. (2004). *Security Analysis*. McGraw-Hill.

Ben-Rephael A., Kandel S., Wohl A. (2010). Measuring Investor Sentiment with Mutual Fund Flows. Forthcoming in the *Journal of Financial Economics*.
http://www.tau.ac.il/~azibenr/JFE_BKW.pdf.

Bertsekas D.P. (1999). *Nonlinear programming*. Athena Scientific, Belmont, MA, USA, 2ième edition.

Bittencouta F.R., Zárata L.E. (2011). Hybrid structure based on previous knowledge and GA to search the ideal neurons quantity for the hidden layer of MLP-Application in the cold rolling process. *Applied Soft Computing*, 11, 2460-2471.

Bliss R.R., Flannery M.J. (2000). Market discipline in the governance of U.S. Bank Holding Companies: monitoring vs. influencing. Working Paper Series WP-00-3, Federal Reserve Bank of Chicago.

Blume L., Easley D., O'Hara m. (1994). Market Statistics and Technical Analysis: The Role of Volume. *The Journal of Finance*, 49 (1), 153-181.

Bollerslev T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *Review of Economics and Statistics*, 69, 542-547.

Box G. E. P., Jenkins G. (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day. San Francisco.

Boyacioglu M.A, Avci D. (2010). An Adaptive Network-Based Fuzzy Inference System (ANFIS) for the prediction of stock market return: The case of the Istanbul Stock Exchange. *Expert Systems with Applications*, 37, 7908-7912.

Brailsford T.J., Faff R.W. (1993). Modelling Australian Stock Market Volatility. *Australian Journal of Management*, 18, 109-132.

Breusch T.S., Godfrey L.G. (1981). A review of recent work on testing for autocorrelation in dynamic simultaneous models. In D.A. Currie, R. Nobay and D. Peels, (eds.), *Macroeconomic Analysis, Essays in Macroeconomics and Economics*, Croom, Helm, London, 63-100.

Brockwell P.J., Davis R.A. (2002). *Introduction to Time Series and Forecasting*. 8th Printing Edition, Springer.

Brown G.W., Cliff M.T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, 11, 1-27.

Brown G.W., Cliff M.T. (2005). Investor sentiment and asset valuation. *Journal of Business*, 78 (2), 405-440.

Brownstone D. (1996). Using percentage accuracy to measure neural network predictions in stock market movements. *Neurocomputing*, 10, 237-250.

Burkitt A.N. (1991). Optimization of The Architecture of Feed-forward Neural Networks with Hidden Layers by Unit Elimination. *Complex Systems*, 5, 371-380.

Burney S.M.A.A , Jilani T.A., Ardil C. (2005). Levenberg-Marquardt Algorithm for Karachi Stock Exchange Share Rates Forecasting. *International Journal of Computational Intelligence*, 1 (3), 144-149.

Campbell S.D., Diebold F.X. (2009). Stock Returns and Expected Business Conditions: Half a Century of Evidence. *Journal of Business and Economic Statistics*, 27, 266-278.

Cano C., García F., López F.J., Blanco A. (2009). Intelligent system for the analysis of microarray data using principal components and estimation of distribution algorithms. *Expert Systems with Applications*, 36, 4654-4663.

Cao L., Tay F.E.H. (2001). Financial Forecasting Using Support Vector Machines. *Neural Comput & Applic*, 10, 184-192.

Carmona R.A. (2003). *Statistical Analysis of Financial Data in S-Plus*. Springer.

Carnero M.A., Peña D., Ruiz E. (2004). Persistence and Kurtosis in GARCH and Stochastic Volatility Models. *Journal of Financial Econometrics*, 2, 319-342.

Cengiz C.-B., Herwartz H. (2011). Modeling stock index returns by means of partial least-squares methods: An out-of-sample analysis for three stock markets. *Appl. Stochastic Models Bus. Ind.*, 27, 253-266.

Chakradhara P., Narasimhan V. (2007). Forecasting exchange rate better with artificial neural network. *Journal of Policy Modeling*, 29, 227-236.

Chang C.-Y, Chung P.-C, Lai P.-H. (2002). Using a Spatiotemporal Neural Network on Dynamic Gadolinium-Enhanced MR Images for Diagnosing Recurrent Nasal Papilloma. *IEEE Transactions on Nuclear Science*, 49 (1), 225 -238.

Chang P.-C., T. Liao W., Lin J.-J., Fan C.-Y. (2011). A dynamic threshold decision system for stock trading signal detection. *Applied Soft Computing*, 11, 3998-4010.

Charniya N.N, Dudul S.V. (2011). Classification of material type and its surface properties using digital signal processing techniques and neural networks. *Applied Soft Computing*, 11,1108-1116.

Charoenrook A. (2003). Change in Consumer Sentiment and Aggregate Stock Market Returns. The Owen Graduate School of Management, Vanderbilt University, Working Paper.

Chen A.S., Leung M.T., Daouk H. (2003). Application of neural networks to an emerging financial market: Forecasting and trading the Taiwan Stock Index. *Computers and Operations Research*, 30, 901-923.

Chen W.-C. (2006). Nonlinear Dynamics and Chaos in Fractional-Order Financial System. *Chaos, Solutions and Fractals*, 36, 1305-1314.

Chen W.-H., Shih J.-Y. (2006). Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets. *Int. J. Electronic Finance*, 1 (1), 2006, 49-67.

Cheng T., Wang J. (2006). Applications of Spatio-Temporal Data Mining and Knowledge for Forest Fire. *ISPRS Technical Commission VII Mid Term Symposium, Enschede*, 148- 53.

Cheng T., Wang J. (2007). Application of a Dynamic Recurrent Neural Network in Spatio-Temporal Forecasting. *International Workshop on Information Fusion and Geographical Information Systems*, May 27-29, St. Petersburg.

Cheung Y., Lai, K. (1995). A Search for Long Memory in International Stock Market Returns. *Journal of International Money and Finance*, 14, 597-615.

Cheung Y.W., Lai K.S. (1995). A Search for Long Memory in International Stock Market Returns. *Journal of International Money and Finance*, 24, 597-615.

Chi L., Zhuang X., Song D. (2011). Investor sentiment in the Chinese stock market: an empirical analysis. *Applied Economics Letters*, 1-4.

Choi J.H., Lee M.K., Rhee M.W. (1995). Trading S&P 500 stock index futures using a neural network. In R. S. Freedman, editor. *Proceedings of the 3rd annual international conference on artificial intelligence applications on Wall Street*, June, New York, 63-72.

Chui C.K. (1992). *An Introduction to Wavelets*. San Diego: Academic Press.

Chung P.-C., Chen E.-L., Wu J.-B. (1998). A Spatiotemporal Neural Network for Recognition Partially Occluded Objects. *IEEE Transactions on Signal Processing*, 46, 1991-2000.

Cipollini A., Kapetanios G. (2009). Forecasting Financial Crises and Contagion in Asia using Dynamic Factor Analysis. *Journal of Empirical Finance*, 16, 188-200.

Coifman R., Meyer Y., Quake S., Wickerhauser V. (1990). Signal processing and compression with wave packets. Numerical Algorithms Research Group, New Haven, CT: Yale University.

Coifman R.R., Wickerhauser M.V. (1990). Best-adapted wave packet bases. Preprint, Yale University.

Coifman R.R., Wickerhauser M.V. (1992). Entropy-Based Algorithms for Best Basis Selection. IEEE Transactions on Information Theory, 38(2), 713-718.

Conrad C., Karanasos M., Zeng N. (2011). Multivariate Fractionally Integrated APARCH Modeling of Stock Market Volatility: A Multi-Country Study. Journal of Empirical Finance, 18, 147-159.

Constantinou E., Georgiades R., Kazandjian A., Kouretas G.P. (2006). Regime switching and artificial neural network forecasting of the Cyprus Stock Exchange daily returns. International Journal of Finance and Economics, 11 (4), 371-383.

Cortes C., V. Vapnik V. (1995). Support vector networks, Machine Learning, 20, 1-25.

Cost S., Salzberg S. (1996). A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features, in: Machine Learning, Morgan Kaufmann, 57-58.

Crato N. (1994). Some International Evidence Regarding The Stochastic Behaviour of Sstock Rreturns. Applied Financial Economics, 4, 33-39.

Cressie N., Majure J.J. (1997). Spatio-Temporal Statistical Modelling of Livestock Waste in Streams. Journal of Agricultural, Biological and Environmental Statistics, 2 (5), 20-28.

Cutler D., Poterba J., Summers L. (1990). Speculative dynamics and role of feedback traders. American Economic Review, 80, 63-68.

Cybenko G. (1989). Approximation by Superpositions of Sigmoidal Function. Math. Contr. Signals Syst., 2, 303-314.

Danial S.N., Noor S.R., Usmani B.A., Zaidi S.J. (2009). A Dynamical System and Neural Network Perspective of Karachi Stock Exchange. Communications in Computer and Information Science, 20, 88-99.

Darrat AF. (1990). Stock Returns, Money and Fiscal Deficits. Journal of Financial and Quantitative Analysis, 25, 387-398.

Daubechies I.(1992). Ten Lectures on Wavelets. SIAM.

Davidson R., MacKinnon J.G. (1993). Estimation and Inference in Econometrics. Oxford University Press.

De Bondt W.F., Tahler R.H. (1985). Does the stock market overreact? Journal of Finance, 40, 793-807.

De Bondt W.F.M. (1993). Betting on trends: Intuitive forecasts of financial risk and return. International Journal of forecasting, 9, 355-371.

De Faria E.L., Albuquerque M.P., Gonzalez J.L., Cavalcante J.T.P. (2009). Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods. Expert Systems with Applications, 36, 12506-12509.

DeLong J.B., Shleifer A., Summers L., Waldmann R.J. (1990). Noise trader risk in financial markets. Journal of Political Economy, 98, 703-738.

Deutsch S.J, Ramos J.A. (1986). Space-Time Modelling of Vector Hydrologic Sequences. Water Resources Bulletin, 22 (6), 967-980.

Devaney R.L. (2003). An Introduction to Chaotic Dynamical Systems, 2nd ed. Westview Press.

Dickey D.A., Fuller W.A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, 427-431.

Doesken B., Abraham A., Thomas J., Paprzycki M. (2005). Real stock trading using soft computing models. In *Proceedings of international symposium on information technology: Coding and computing ITCC*. 2, 162-167.

Dong I., Duan C., Jang M.-J. (2003). Predicting extreme stock performance more accurately. A paper written for "Government 2001".

Dong M., Zhou X. (2002). Exploring the fuzzy nature of technical patterns of U.S. stock market. *Proceedings of Fuzzy System and Knowledge Discovery*, 1, 324-328.

Dong M., Zhou X.-S. (2008). Knowledge discovery in corporate events by neural network rule extraction. *Appl Intell*, 29, 129-137.

Dourra H., Siy P. (2002). Investment using technical analysis and fuzzy logic. *Fuzzy Sets and Systems*, 127, 221-240.

Dunis C.L., Zhou B. (1998). *Nonlinear Modelling of High Frequency Financial Time Series*. Wiley Publishing.

Edwards F.R., Zhang X. (1998). Mutual funds and stock and bond market stability. *Journal of Financial Services Research*, 13, 257-282.

Egeli B., Ozturan M., Badur B. (2003). Stock market prediction using artificial neural networks. *Proceedings of 3rd Hawaii International Conference on Business*.

Eleftherios G. (2010). Application of logit model and self-organizing maps (SOMs) for the prediction of financial crisis periods in US economy. *Journal of Financial Economic Policy*, 2 (2), 98-125.

Eligiüs M.T., Boglárka G.-T. (2010). *Introduction to Nonlinear and Global Optimization*. Springer, 1st Edition. Edition.

Ellul A. (2006). Ripples Through Markets: Inter-Market Impacts Generated by Large Trades. *Journal of Financial Economics*, 82, 173-196.

Embrechts M., Cader M., Deboeck G.J. (1994). Nolinear Dimensions of Foreign Exchange, Stock and Bond Markets. in *Trading on The Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*, G. J. Deboeck, Eds. New York: Wiley, 297-313.

Engle R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation. *Econometrica*, 50, 987-1008.

Engle R.F., Rangel J.G. (2008). The Spline-GARCH Model for Low-Frequency Volatility and its Global Macroeconomic Causes. *Review of Financial Studies*, 21, 1187-1222.

Enke D., Thawornwong S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications*, 29, 927-940.

Esfahanipour A., Aghamiri W. (2010). Adapted Neuro-Fuzzy Inference System on indirect approach TSK fuzzy rule base for stock market analysis. *Expert Systems with Applications*, 37, 4742-4748.

Eshaghi M., Mollaei M.R.K. (2010). Voice activity detection based on using wavelet packet. *Digital Signal Processing*, 20, 1102-1115.

Fabozzi F.J. (2001). *Handbook of Portfolio Management*. McGraw Hill.

Fama E.F., French K.R. (2002). The Equity Premium. *Journal of Finance*, 57, 637-59.

Fant L.F. (1999). Investment behavior of mutual fund shareholders: The evidence from aggregate fund flows. *Journal of Financial Markets*, 2, 391-402.

Fausett L. (1994). *Fundamentals of Neural Networks, Architectures, Algorithms, and Applications*, Prentice-Hall, Inc., Upper Saddle River, NJ.

Fernandez-Rodriguez F., Gonzalez-Martel C., Sosvilla-Rivebo S. (2000). On the profitability of technical trading rules based on artificial neural networks: Evidence from the Madrid Stock Market. *Economics Letters*, 69, 89-94.

Fisher K.L., Statman M. (2000). Investor Sentiment and Stock Returns. *Financial Analysts Journal*, 56 (2), 16-23.

Fisher K.L., Statman M. (2003). Consumer Confidence and Stock Returns. *Journal of Portfolio Management*, 30 (1), 115-128.

Flannery M.J., Protopapadakis A.A. (2002). Macroeconomic Factors Do Influence Aggregate Stock Returns. *The Review of Financial Studies*, 15, 751-82.

Funahashi K.-I. (1989). On the Approximate Realization of Continuous Mappings by Neural Networks. *Neural Networks*, 2, 183-192.

Gang G, Zhongke S., Guanzhong D. (2000). Stock Market Multi-step Forecast Using Chaos Theory. *Information and Control*, (4), 77-81.

Garliauskas, A. (1999). Neural Network Chaos and Computational Algorithms of Forecast in Finance. *IEEE International Conference on Systems, Man, and Cybernetics*, 2, 638-643.

Gençay R., Selcuk F., Whitcher B. (2001). Scaling properties of foreign exchange volatility. *Physica* 289, 249-66.

Geweke J., Porter-Hudak S. (1983). The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis*, 4, 221-38.

Girish K.J., Parimala T., Rupp K.T. (2009). PSO Based Neural Network for Time Series Forecasting. *IEEE Proceedings of International Joint Conference on Neural Networks*, Atlanta, Georgia, USA, June 14-19, 1422-1427.

Golovachev S. (2011). Forecasting the U.S. Stock Market via Levenberg-Marquardt and Haken Artificial Neural Networks Using ICA&PCA Pre-processing Techniques. LNCS 6744, 351-357.

Goyal A., Santa-Clara P. (2003). Idiosyncratic Risk Matters!. *Journal of Finance*, 58, 975-1008.

Goyal A., Welch I. (2003). Predicting The Equity Premium with Dividend Ratios. *Management Science*, 49, 639-54.

Goyal A., Welch I. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies*, 21 (4), 1455-1508.

Granger C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37 (3), 424-438.

Granger C.W.J., Hyung N. (2004). Occasional structural breaks and long-memory. *Journal of Empirical Finance*, 11, 399-421.

Granger C.W.J., Joyeux R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1: 15-30.

Grosan C., Abraham A. (2006). Stock Market Modeling Using Genetic Programming Ensembles. *Studies in Computational Intelligence (SCI)*, 13, 131-146.

Guégan D. (2009). Chaos in Economics and Finance. *Annual Reviews in Control*, 33, 89-93.

Guresen E., Kayakutlu G. (2011). Definition of artificial neural networks with comparison to other networks. *Procedia Computer Science*, 3, 426-433.

Halliday R. (2004). Equity trend prediction with neural networks. *Research Letters in the Information and Mathematical Sciences*, 6, 135-149.

Hamilton J.D. (1989). A new approach to the economic analysis of non-stationary time series and the business cycle. *Econometrica* 57, 357-384.

Hamilton J.D. (1994). *Time Series Analysis*, Princeton University Press, New Jersey.

Hardouvelis G.A. (1987). Macroeconomic information and stock prices. *Journal of Economics and Business*, Elsevier, 39 (2), 131-140.

Harris R., Sollis R. (2010). *Applied Time Series Modelling and Forecasting*. Wiley.

Hartwig J. (2010). Is Health Capital Formation Good for Long-Term Economic Growth? – Panel Granger-Causality Evidence for OECD Countries. *Journal of Macroeconomics*, 32, 314-325.

Harvey C.R. (2001). The Specification of Conditional Expectations. *Journal of Empirical Finance*, 8, 573-637.

Harvey C.R., Travens K.E., Costa M.J. (2000). Forecasting emerging market returns using neural networks. *Emerging Markets Quarterly*, 4 (2), 43-55.

Haykin S. (2001). *Intelligent Signal Processing*. Wiley-IEEE Press.

He K., Xie C., Shou Chen, Lai K.K. (2009). Estimating VaR in crude oil market: A novel multi-scale non-linear ensemble approach incorporating wavelet analysis and neural network. *Neurocomputing*, 72, 3428-3438.

Hillebrand E., Medeiros M.C. (2006). Estimating and Forecasting GARCH Models in The Presence of Structural Breaks and Regime Switches. In: Rapach, D. and Wohar, M, (eds.) *Forecasting in the Presence of Structural Breaks and Uncertainty*. Frontiers of Economics and Globalization, Elsevier.

Hirshleifer D. (2001). Investor Psychology and Asset Pricing. *The Journal of Finance*, 56 (4), 1533-1597.

Hirshleifer D.K., Subrahmanyam A. (1997). A theory of overconfidence, self-attribution, and security market under- and over-reactions. Unpublished working paper. University of Michigan.

Ho C., Hung C.-H. (2009). Investor Sentiment as Conditioning Information in Asset Pricing. *Journal of Banking and Finance*, 33(5), 892-903.

Hornik K. (1991). Approximation Capabilities of Multilayer Feedforward Networks. *Neural Networks*, 4, 251-257.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, 68, 165-176.

Hribar P., McNinnis J. (2011). Investor sentiment and analyst's forecast errors. Social Science Research Network. <http://ssrn.com/abstract=1452893>.

Hsieh T.-J., Hsiao H.-F., Yeh W.-C. (2011). Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm. *Applied Soft Computing*, 11, 2510-2525.

Hu C., Xie K., Song G., Wu T. (2008). Hybrid Process Neural Network based on Spatio-Temporal Similarities for Short-Term Traffic Flow Prediction. *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems* Beijing, China, October 12-15, 253-258.

Huang C.-F., Chang B.R., Cheng D.-W., Chang C.-H. (2012). Feature Selection and Parameter Optimization of a Fuzzy-based Stock Selection Model Using Genetic Algorithms. *International Journal of Fuzzy Systems*, 14 (1), 65-75.

Huang S.-C. (2011). Forecasting stock indices with wavelet domain kernel partial least square regressions. *Appl. Soft Comput. J.*, doi:10.1016/j.asoc.2011.05.015.

Huang S.-C., Wu T.-K. (2010). Integrating recurrent SOM with wavelet-based kernel partial least square regressions for financial forecasting. *Expert Systems with Applications*, 37, 5698-5705.

Huang W., Lai K., Nakamori Y., Wang S.Y., Yu L. (2007). Neural networks in finance and economics forecasting. *International Journal of Information Technology and Decision Making*, 6, 113-140.

Huang W., Nakamori Y., Wang S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computer and Operations Research*, 32, 2513-2522.

Jaruszewicz M., Mandziuk J. (2004). One day prediction of NIKKEI Index considering information from other stock markets. *Lecture Notes in Computer Science* (3070). Springer.

Jensen A., la Cour-Harbo A. (2001). *Ripples in Mathematics: The Discrete Wavelet Transform*. Springer Verlag.

Jinjarak Y., Sheffrin S.M. (2011). Causality, Real Estate Prices, and The Current Account. *Journal of Macroeconomics*, 33, 233-246.

John G., Kohavi R., Pfleger K. (1994). Irrelevant Features and The Subset Selection Problem, in: *Proceeding of the 11th International Conference on Machine Learning*, Morgan Kaufmann, San Francisco, USA, 121-129.

Kahneman D., Tversky A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-292.

Kahneman D.E., Riepe M.W. (1998). Aspects of investor psychology. *Journal of Portfolio Management*, 24, 52-67.

Kamruzzaman J., Sarker R.A. (2003). Comparing ANN Based Models with ARIMA for Prediction of Forex Rates. *ASOR Bulletin*, 1-10.

Kara Y., Boyacioglu M.A., Baykan Ö.K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, 38, 5311–5319.

Kavzoglu T. (1999). Determining Optimum Structure for Artificial Neural Networks. In *Proceedings of the 25th Annual Technical Conference and Exhibition of the Remote Sensing Society*, Cardiff, UK, 8-10 September, 675-682.

Kayse A.S, Sun F.T, D'Esposito M. (2009). A Comparison of Granger Causality and Coherency in fMRI-Based Analysis of The Motor System. *Human Brain Mapping*, 30 (11), 3475-3494.

Khashei M., Bijari M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications*, 37 (1), 479-489.

Kim K.-J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55, 307- 319.

Kim K.-J. (2006). Artificial neural networks with evolutionary instance selection for financial forecasting. *Expert System with Applications*, 30, 519-526.

Kohavi R., Sommerfield D. (1995). Feature Subset Selection using The Wrapper Method: Overfitting and Dynamic Search Space Topology, in: *The First International Conference on Knowledge Discovery and Data Mining*.

Koller D., M. Sahami. (1996). Toward Optimal Feature Selection. In Lorenza Saitta, ed., *Machine Learning: Proc. of the Thirteenth International Conference*, Morgan Kaufmann.

Koopman S.J., Jungbacker B., Hol, E. (2005). Forecasting Daily Variability of The S&P 100 Stock Index using Historical, Realised and Implied Volatility Measurements. *Journal of Empirical Finance*, 12, 445-475.

Kubinyi M., Kreibich O., Neuzil J., Smid R. (2011). EMAT Noise Suppression using Information Fusion in Stationary Wavelet Packets. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 58 (5), 1027-1036.

Kumar A., Lee M.C. (2006). Retail Investor Sentiment and Return Comovements. *The Journal of Finance*, 61 (5), 2451-2486.

Lai L.K.C., Liu J.N.K. (2010). Stock forecasting using support vector machine. *Proceedings of the Ninth International Conference on Machine Learning and Cybernetics*, Qingdao, 11-14 July 2010, 1607-1614.

Lam S. S. (2001). A genetic fuzzy expert system for stock market timing. *Proceedings of the IEEE Conference on Evolutionary Computation*, 410-417.

Lawrence M.J., Makridakis S. (1989). Factors affecting judgmental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes*, 42, 172-187.

Leigh W., Paz M., Purvis R. (2002). An analysis of a hybrid neural network and pattern recognition technique for predicting short-term increases in the NYSE Composite Index. *Omega*, 30, 69-76.

Lemmon M., Portniaguina E. (2006). Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies*, 19, 1499-1529.

Lendasse A., De Bodt E., Wertz V., Verleysen M. (2000). Non-linear financial time series forecasting – Application to the Belgium 20 Stock Market Index. *European Journal of Economical and Social Systems*, 14 (1), 81-91.

Lettau M., Ludvigson S. (2010). Measuring and Modeling Variation in The Risk-Return Trade-Off. In: *Handbook of Financial Econometrics*, 1, 617-690.

Levenberg. K. (1944). A Method for the Solution of Certain Problems in Least Squares. *Quart. Appl. Math.* 2, 164-168.

Levich R., Thomas L. (1993). The Significance of Technical Trading Rule Profits in the Foreign Exchange Market: A Bootstrap Approach. *Journal of International Money and Finance*, 12, 451-474.

Li F. (2010). Research on the Estimating Model of the Stock Market Price Based on the LM-BP Neural Network. *ICGEC, Fourth International Conference on Genetic and Evolutionary Computing*, 562-565.

Li J., Shi Z., Li X. (2006). Genetic Programming with Wavelet-Based Indicators for Financial Forecasting. *Transactions of the Institute of Measurement and Control*, 28 (3), 285-297.

Lichtenstein S., Fischhoff B., Phillips L.D. (1982). Calibration of Probabilities: The State of the Art to 1980. In *Judgement under Uncertainty: Heuristics and Biases*, ed. D. Kahneman and A. Tversky. Cambridge University Press.

Lieberman O., Phillips P.C.B. (2008). Refined Inference on Long Memory in Realized Volatility. *Econometric Reviews*, 27, 254-267.

Lintner J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47 (1), 13-37.

Liu H., Motoda H. (1998). *Feature Selection for Knowledge Discovery and Data Mining*. Boston: Kluwer Academic.

Liu H., Yu L. (2005). Toward Integrating Feature Selection Algorithms for Classification and Clustering. *IEEE Trans. On Knowledge and Data Engineering*, 17 (4), 491-502.

Liu J., Khalil H.K, Oweiss K.G. (2011). Neural Feedback for Instantaneous Spatio-Temporal Modulation of Afferent Pathways in Bi-Directional Brain-Machine Interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19 (5), 521-533.

Ljung G.M, Box G.E.P. (1978). On a Measure of a Lack of Fit in Time Series Models. *Biometrika*, 65 (2), 297-303.

Lo A.W. (1991). Long-Term Memory in Stock Market Prices, *Econometrica*. 59, 1279- 1313.

Lo A.W., MacKinlay A.C. (1988). Stock Market Prices Do Not Follow Random Walks: Evidence From A Simple Specification Test. *Review of Financial Studies*, 1, 41-66.

Ludvigson S., Ng S. (2007). The Empirical Risk-Return Trade-Off: A Factor Analysis Approach. *Journal of Financial Economics*, 83, 171-222.

Lux T. (2010). Sentiment Dynamics and Stock Returns: The Case of The German Stock Market. *Empirical Economics*, 41, 663-679.

Lux T. (2012). Estimation of An Agent-Based Model of Investor Sentiment Formation in Financial Markets. *Journal of Economic Dynamics & Control*, 36, 1284-1302.

Majhi R., Panda G. (2007). Stock market prediction of S&P 500 and DJIA using bacterial foraging optimization technique. *IEEE congress on evolutionary computation (CEC 2007)*, 2569-2579.

Majhi R., Panda G., Majhi B., Sahoo G. (2009). Efficient prediction of stock market indices using adaptive bacterial foraging optimization (ABFO) and BFO based techniques. *Expert Systems with Applications*, 36, 10097-10104.

Malkiel B., Xu Y. (2003). Investigation the Behavior of Idiosyncratic Volatility. *Journal of Business*, 76, 613-644.

Mallat S.G. (1989). A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7 (11), 674-693.

Mallat S.G. (1999). A wavelet tour of signal processing. Academic Press.

March J., Shapira Z. (1992). Variable risk preferences and the focus of attention. *Psychological Review*, 99, 172-183. Reprinted in J. March (Ed) *The pursuit of intelligence in organizations*, Cambridge, Mass.: Blackwell, 1999.

Marquardt D. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM J. Appl. Math.* 11, 431-441.

Marquering W., Verbeek M. (2004). The Economic Value of Predicting Stock Index Returns and Volatility. *Journal of Financial and Quantitative Analysis*, 39, 407-429.

McKenzie M.D. (2001). Non-Periodic Australian Stock Market Cycles: Evidence from Rescaled Range Analysis. *The Economic Record*, 77 (239), 393-406.

Mills T. (1993). Is there long-term memory in UK stock returns?. *Applied Financial Economics*, 3, 303-306.

Mizuno H., Kosaka M., Yajima H. (1998). Application of neural network to technical analysis of stock market prediction. *Studies in Informatics and Control*, 7, 111-120.

Mochón A., Quintana D., Sáez Y., Isasi P. (2008). Soft computing techniques applied to finance. *Appl Intell*, 29, 111-115.

Moeini A., Ahrari M., Karimi P. (2010). Forecasting Gold Price via Chaotic Models and Lyapunov Exponent. *Middle Eastern Finance and Economics*, 8, 79-93.

Mossin J. (1966). Equilibrium in a Capital Asset Market. *Econometrica*. 34 (4), 768-783.

Motiwalla L., Wahab M. (2000). Predictable variation and profitable trading of US equities: A trading simulation using neural networks. *Computer and Operations Research*, 27, 1111-1129.

Mukherjee C.S., Sarkar A. (2011). Long Memory in Stock Returns: Insights from the Indian Market. *The International Journal of Applied Economics and Finance*, 5, 62-74.

Müller K.R., Mika S., Rätsch G., Tsuda K., Schölkopf B. (2001). An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Networks*, 12 (2), 181-201.

Murtagh F., Starck J.L., Renaud O. (2003). On neuro-wavelet modeling. *The Journal of Decision Support System*, 37, 475-484.

Neal R., Wheatley S.M. (1998). Do Measures of Investor Sentiment Predict Returns?. *Journal of Financial and Quantitative Analysis*, 33 (4), 523-547.

Neely C., Weller P., Dittmar R. Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach. (1997). *The Journal of Financial and Quantitative Analysis*, 32 (4), 405-426.

Nelson D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59, 347-370.

Newey W.K, West K.D. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55 (3), 703-708.

Nocedal J., Wright S.J. (2006). *Numerical Optimization*. Springer.

Noori R., Khakpour A., Omidvar B., Farokhnia A. (2010). Comparison of ANN and principal component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistic. *Expert Systems with Applications*, 37, 5856-5862.

Nyberg H. (2011). Forecasting the direction of the US stock market with dynamic binary probit models. *International Journal of Forecasting*, 27, 561-578.

Odean T. (1998). Are investors reluctant to realize their losses. *Journal of Finance*, 53 (5), 1775-1798.

Odean T. (1998). Volume, Volatility, Price, and Profit When All Traders Are above Average. *Journal of Finance*, 53, 1887-1934.

Oh K.J., Kim K. (2002). Analyzing Stock Market Tick Data using Piecewise Nonlinear Model. *Expert Systems with Applications*, 22, 249-255.

Panchal G., Ganatra A., Kosta Y.P, Panchal D. (2011). Behaviour Analysis of Multilayer Perceptrons with Multiple Hidden Neurons and Hidden Layers. *International Journal of Computer Theory and Engineering*, 3 (2), 332-337.

Papagiannaki K., Taft N., Zhang Z.-L., Diot C. (2005). Long-term forecasting of internet backbone traffic, *IEEE Transactions on Neural Networks*, 1110-1124.

Percival D.B., Andrew T.W. (2000). *Wavelet Methods for Time Series Analysis* by Percival and Walden, Cambridge University Press.

Pfeifer P.E, Deutsch S.J. (1990). A Statima Model-Building Procedure with Application to Description and Regional Forecasting. *Journal of Forecasting*, 9, 50-59.

Pokrajac D., Obradovic Z. (2001). Improved Spatial-Temporal Forecasting Through Modelling of Spatial Residuals in Recent History. *Proceedings of first international SIAM conference on data-mining*, Chicago, 368-386.

Porterba J., Summers L.H. (1988) Mean reversion in stock prices: evidence and implications. *Journal of Financial Economics*, 22, 27-59.

Powell M.J.D. (1977). Restart procedures for the conjugate gradient method. *Mathematical Programming*, 12, 241-254.

Qi M., Zhang G.P. (2001). An Investigation of Model Selection Criteria for Neural Network Time Series Forecasting. *European Journal of Operational Research*, 132, 666-680.

Qian B., Rasheed K. (2004). Hurst exponent and financial market predictability. In: *Proceedings of The 2nd IASTED international conference on financial engineering and applications*. Cambridge, MA, USA, 203-209.

Qian B., Rasheed K. (2007). Stock market prediction with multiple classifiers. *Applied Intelligence*, 26, 25-33.

Qiu L., Welch I. (2006). Investor sentiment measures. Working paper. Brown University.

Ramchandran K., Vetterli M. (1993). Best wavelet packet bases in a rate-distortion sense. *IEEE Trans. Image Processing*, 2, 160-175.

Ramsey J.B. (1999). The contribution of wavelets to the analysis of economic and financial data, *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, 2593-2606.

Rawat M., Ghannouchi F.M. (2012). Distributed Spatiotemporal Neural Network for Nonlinear Dynamic Transmitter Modeling and Adaptive Digital Predistortion. *IEEE Transactions on Instrumentation and Measurement*, 61 (3), 595-608.

Rege S., Martín S.G. (2011). Portuguese stock market: A long-memory process ?. *Business Theory and Practice*, 12 (1), 75-84.

Reilly F.K., Brown C.K. (1997). *Investment Analysis and Portfolio Management*. Sixth Edition. Fort Worth: The Dryden Press.

Robert N., Simon M. Wheatley S.M. (1998). Do Measures of Investor Sentiment Predict Returns?. *Journal of Financial & Quantitative Analysis*, 33 (4), 523-548.

Robinson P.M. (2003). *Time Series With Long Memory*, Oxford University Press.

Ruiz N., Rosa M., López F., Martínez D. (1999). An adaptive wavelet-based approach for perceptual low bit rate audio coding attending to entropy-type criteria, *Software and Hardware Engineering for the 21th Century*, WSES Press.

Scales L.E. (1985). *Introduction to Non-Linear Optimization*. New York, Springer-Verlag.

Schmeling M. (2007). Institutional and individual sentiment: Smart money and noise trader risk?. *Int. J. Forecast.*, 23, 127-145.

Scholkopf B., Mika S., Burges J.C., Knirsch P., Muller K.-R., Ratsch G., Smola A. (1999). Input space versus feature space in Kernel-based methods. *IEEE Trans. On Neural Networks*, 10 (5), 1000-1017.

Scholkopf B., Smola A.J. (2002). Learning with Kernels: Support Vector Machines, Regularization and Beyond. MIT Press.

Scholkopf B., Smola A.J., Williamson R.C., Bartlett P.L. (2000). New support vector algorithms. Neural Computation, 1207-1245.

Schwager J.D. (1999). Getting Started in Technical Analysis. Wiley.

Schwert G.W. (1990). Stock Returns and Real Activity: A Century of Evidence. NBER Working Papers 3296, National Bureau of Economic Research.

Shafi I., Ahmad J., Shah S.I., Kashif F.M. (2006). Impact of Varying Neurons and Hidden Layers in Neural Network Architecture for a Time Frequency Application. IEEE Multitopic Conference, 188-193.

Shannon C.E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379-423.

Shapira Z., Venezia I. (2000). Patterns of behavior in professionally managed and independent investors. Journal of Banking & Finance, 25, 1573-1588.

Sharma A.K, Sharma R.K., Kasana H.S. (2007). Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. *Applied Soft Computing*, 7, 1112-1120.

Sharpe W.F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19 (3), 425-442.

Shefrin H. (2002). *Beyond Greed and Fear: Understanding behavioral finance and the psychology of investing*. New York: Oxford University Press.

Shefrin H., Statman M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance*, 40, 777-790.

Shibata K. (2009). Effect of number of hidden neurons on learning in large-scale layered neural networks. ICROS-SICE International Joint Conference, Fukuoka International Congress Center, Japan, August 18-21, 5008-5013.

Shui P.L., Zhou Z.F., Li J. X. (2007). Image Denoising Algorithm via Best Wavelet Packet Base using Wiener Cost Function. *IET Image Processing*, 1 (3), 311-318.

Simon D.P., Wiggins R.A. (2001). S&P Futures Returns and Contrary Sentiment Indicators. *Journal of Futures Markets*, 21, 447-462.

Simon G., Odean T. (2001). Learning to Be Overconfident. *Review of Financial Studies*, 14, 1-27.

Simon H. (1987). *Behavioral Economics*. The New Palgrave: A Dictionary of Economics, 1, 221-24.

Smit G.F, Jordaan E.M. (2002). Improved SVM Regression using Mixtures of Kernels. *IEEE International Joint Conference on Neural Networks*, 3, 2785-2790.

Sprott J.C. (2003). *Chaos and Time-Series Analysis*. Oxford University Press.

Strang G., Nguyen T. (1996). *Wavelets and Filter Banks*. Wellesley: Wellesley-Cambridge Press.

Tan T.Z., Quek C., Ng G.S. (2005). Brain inspired genetic complimentary learning for stock market prediction. In *Proceedings of the IEEE congress on evolutionary computation*, 2-5 September, 3, 2653-2660.

Tan Z., Quek C., Cheng P.Y.K. (2011). Stock trading with cycles: A financial application of ANFIS and reinforcement learning. *Expert Systems with Applications*, 38, 4741-4755.

Tay F.E.H., Cao L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 29, 309-317.

Tay F.E.H., Cao L.J. (2001). Improved financial time series forecasting by combining Support Vector Machines with self-organizing feature map. *Intelligent Data Analysis*, 5 339-354.

Taylor M.P., Allen H. (1992). The Use of Technical Analysis in the Foreign Exchange Market. *Journal of International Money and Finance*, 11, 304-314.

Terence C. Mills T.C. (1999). *The Econometric Modelling of Financial Time Series*. Second edition, Cambridge University Press.

Tetlock P.C. (2007). Giving Content to Investor Sentiment: The Role of Media in The Stock Market. *The Journal of Finance*, 62 (3), 1139-1168.

Thawornwong S., Enke D. (2004). The adaptive selection of financial and economic variables for use with artificial neural networks. *Neurocomputing*, 56, 205-232.

Tsaih R., Hsu Y., Lai C.C. (1998). Forecasting S&P 500 stock index futures with a hybrid AI system. *Decis Support System*, 23 (2), 161-174.

Tsibouris G., Zeidenberg M. (1995). Testing the efficient markets hypothesis with gradient descent algorithms, In: Refenes AP (ed) *Neural networks in the capital markets*. JohnWiley & Sons, Chichester, England, Chap 8, 127-136.

Turtle H.J, Zhang C. (2012). Time-Varying Performance of International Mutual Funds. *Journal of Empirical Finance*, 19, 334-348.

Unnikrishnan K.P, Hopfield J.J, Tank D.W. (1991). Connected-Digit Speaker-Dependent Speech Recognition using a Neural Network with Time-Delayed Connected. *IEEE Transactions on Signal Processing*, vol. 39, pp.698-713, 1991.

Vapnik V.N. (1995). *The Nature of Statistical Learning Theory*, 2nd edition, New York: Springer-Verlag.

Verma R., Baklaci H., Soyдеми G. (2008). The impact of rational and irrational sentiments of individual and institutional investors on DJIA and S&P500 index returns. *Applied Financial Economics*, 18 (1), 1-15.

Verma R., Verma P. (2007). Noise trading and stock market volatility. *J. Mult. Financial Mgmt.*, 17, 231-243.

Walczak S. (2001). An empirical analysis of data requirements for financial forecasting with neural networks. *Journal of Management Information Systems*, 17, 203-222.

Wang C. (2001). Investor sentiment and return predictability in agricultural futures markets. *Journal of Futures Markets*, 21 (10), 929-952.

Wang C. (2003). Investor sentiment, market timing, and futures returns. *Applied Financial Economics*, 13 (12), 891-898.

Wang C.-M., Huang Y.-F. (2009). Evolutionary-based feature selection approaches with new criteria for data mining: A case study of credit approval data. *Expert Systems with Applications*, 36, 5900-5908.

Wang J.H, Leu J.Y. (1996). Stock market trend prediction using ARIMA-based neural networks. *Proceedings of IEEE international Conference on Neural Networks*, 2160-2165.

Wang J.-Z., Wang J.-J., Zhang Z.-G., Guo S.-P. (2011). Forecasting stock indices with back propagation neural network. *Expert Systems with Applications* 38, 14346-14355.

Wang Y.F. (2002). Predicting stock price using fuzzy grey prediction system. *Expert Systems with Applications*, 22, 33-39.

Wang Y.H., Keswani A. Taylor S.J. (2006). The relationship between sentiment, returns, and volatility. *Int. J. Forecast.*, 22, 109-123.

Warther V.A. (1995). Aggregate mutual fund flows and security returns. *Journal of Financial Economics*, 39, 209-235.

Whaley R.E. (2000). The Investor Fear Gauge. *Journal of Portfolio Management*, 26 (3), 12-17.

Wiesenberger A. (1946). *Investment Companies Services*. New York, NY: Warren, Gorham, and Lamont, (1943-1991).

Wooldridge J. (2002). *Introductory Econometrics: A Modern Approach*. 2nd Revised edition, South-Western College Publishing.

Wu J.-D., Liu C.-H. (2009). An expert system for fault diagnosis in internal combustion engines using WPT and NN. *Expert systems with applications*, 36 (3), 4278-4286.

Wu X., Fung M., Flitman A. (2001). Forecasting stock market performance using hybrid intelligent system. *Proceedings of the International Conference on Computational Science*, 447-458.

Yang A.S., Wu M.-L. (2011). Exploring the relationship between investor sentiment and price volatility. *Quantitative Finance*, 11 (6), 955-965.

Ye Z., Mohamadian H., Ye Y. (2009). Quantitative Effects of Discrete Wavelet Transforms and Wavelet Packets on Aerial Digital Image Denoising. *IEEE Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, 1-5.

Yiwen Y., Guizhong L., Zongping Z. (2000). Stock market trend prediction based on neural networks. *Multiresolution Analysis and Dynamical Reconstruction*. In *Proceedings of the IEEE/IAFE/INFORMS 2000 Conference on Computational Intelligence for Financial Engineering*, 155-156.

Yu J., Yuan Y. (2011). Investor sentiment and the mean-variance relation. *Journal of Financial Economics*, 100, 367-381.

Yu L., Wang S., Lai K.K. (2005). A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. *Computers & Operations Research*, 32, 2523-2541.

Yu L., Wang S., Lai K.K. (2009). A neural-network-based nonlinear metamodeling approach to financial time series forecasting. *Applied Soft Computing*, 9, 563-574.

Yudong Z., Lenan W. (2009). Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Systems with Applications*, 36, 8849-8854.

Yumlu S., Gurgen F.G., Okay N. (2005). A Comparison of global, recurrent and smoothed-piecewise neural models for Istanbul Stock Exchange prediction. *Pattern Recognition Letters*, 26, 2093-2103.

Yvan S., Inza I., Larranaga P. (2007). A Review of Feature Selection Techniques in Bioinformatics. *Bioinformatics*, 23 (19), 2507-2517.

Zhang D., Jiang Q., Li X. (2004). Application of neural networks in financial data mining. *Proceedings of International Conference on Computational Intelligence*, 392-395.

Zhang G., Peter B., Patuwo E., Hu M.Y. (1998). Forecasting with Artificial Neural Networks: The State of the Art. *International Journal of Forecasting*, 14 (1), 35-62.

Zhang Y., Dong Z., Wu L., Wang S. (2011). A Hybrid Method for MRI Brain Image Classification. *Expert Systems with Applications*, 38, 10049-10053.

Zhang Y., Wang S., Wu L. (2010). A Novel Method for Magnetic Resonance Brain Images Classification Based on Adaptive Chaotic PSO. *Progress In Electromagnetics Research*, 109, 325-343.

Zhao X., Ye B. (2010). Convolution wavelet packet transform and its applications to signal Processing. *Digital Signal Processing*, 20, 1352-1364.

Zhou M.-R, Wu D., Zhang H.-Q, Song C. (2011). Wavelet Packet Transform in Optical Fiber Gas Sensor System of Signal Denoising. *IEEE International Conference on Electric Information and Control Engineering (ICEICE)*, 2843-2846.